Introdução à Álgebra Linear e suas Aplicações

Bruno Santiago

19 de junho de 2019

Sumário

1 Introdução **9**

1.1 O Teorema de Sylvester-Gallai 10

1.2 *Netflix chalenge* 13

2 Espaços vetoriais **15**

2.1 Introdução 15

2.2 Espaços Euclideanos 16

2.2.1 Exemplos de aplicações 17

2.2.2 Espaços euclideanos são exemplos de espaços vetoriais 19

2.3 Outros exemplos de espaços vetoriais 20

2.4 Combinações lineares 23

2.4.1 Algumas combinações lineares especiais 24

2.5 Julia 24

2.5.1 Instalação 25

2.5.2 Usando Julia 25

3 Produto interno **29**

3.1 Introdução 29

3.2 Produto interno euclideano 30
3.2.1 Alguns casos especiais 30
3.2.2 Aplicações 31
3.3 Outros exemplos de produto interno 33
3.3.1 Produto interno euclideano com pesos 34
3.3.2 Integral como produto interno 34

4 Comprimento e ângulo **35**
4.1 Norma 35
4.1.1 A norma do máximo 37
4.1.2 A norma da soma (ou a distância do taxista) 38
4.2 Normas advindas de um produto interno 38
4.2.1 Cauchy-Schwartz 39
4.2.2 Ângulo entre vetores 41
4.3 Distância entre vetores 41

5 Aplicações da norma euclideana **45**
5.1 O valor rms 45
5.1.1 O produto interno rms 46
5.2 A desigualdade de Chebyshev 47
5.3 Medindo distâncias em $\mathbb{R}^d$ 48
5.3.1 Aplicações 48
5.4 Desvio Padrão 49
5.4.1 A média e o vetor centrado 49

5.4.2 O desvio padrão 50

5.4.3 Coeficiente de correlação 56

5.5 Julia 59

5.5.1 A super calculadora Julia 59

6 Aglutinação de dados (*Clustering data*) 61

6.1 Formalização do problema 61

6.1.1 Representantes do grupo 62

6.1.2 Enunciado do problema 63

6.2 Motivação 64

6.3 O algorítimo $k - means$ 64

6.3.1 O passo 1 65

6.3.2 Passo 2 66

6.3.3 Aplicações 68

7 Base e dimensão 75

7.1 Subespaços 75

7.2 Revisando combinações lineares 77

7.2.1 Subespaço gerado 79

7.3 Dependência Linear; Base e Dimensão 80

7.4 Bases 85

7.4.1 Espaços vetoriais de dimensão finita 88

8 Bases ortonormais 91

8.1 O algorítimo de Gram-Schmidt 91

8.2 O complemento ortogonal 93

9 Funções lineares e suas matrizes **97**

9.1 Funções Lineares 98

9.1.1 Matrizes como transformações lineares 100

9.2 A matriz de uma transformação linear 105

9.2.1 Combinações lineares 105

9.2.2 A matriz de f relativa a bases de E e de F 106

9.2.3 Exemplos de aplicações 109

9.3 Funções afins 110

10 Exemplos e aplicações de matrizes **111**

10.1 Aplicações 111

10.2 Tipos especiais de matrizes 112

10.3 Algumas transformações geométricas 113

10.4 Álgebra de matrizes 114

10.4.1 Medindo o comprimento de uma matriz 114

11 Grafos e matrizes I **115**

11.1 A matriz de adjacência de um grafo 117

11.2 A matriz de incidência de um grafo 117

11.3 A energia de Dirichlet de um grafo 118

12 Núcleo e Imagem **119**

12.1 Definição 119

13 Funcionais lineares e aplicações **123**

13.1 Programação Linear 123

14 A transposta **127**

14.1 A matriz transposta 128

14.2 Transposição e ortogonalidade 129

15 Equações Lineares **131**

15.1 O problema do fluxo de tráfego 132

15.2 Coeficientes da combinação linear 135

15.3 O Método da eliminação de Gauss 137

15.3.1 Solução de equações lineares 137

15.3.2 Dimensão do subespaço gerado 138

15.3.3 Cálculo do posto e da nulidade de uma matriz 139

15.4 Interpolação polinomial 141

15.4.1 Implementação em Julia 142

16 Sistemas Dinâmicos Lineares **145**

17 Quociente e Dualidade **147**

18 Produto de matrizes **157**

18.1 Caminhos num grafo 158

18.2 Composição de funções lineares 161

19 Matriz inversa **163**

20 Determinantes: significado e propriedades **165**

21 O método dos mínimos quadrados **167**

22 Autovalores e autovetores **169**

23 Potência de matrizes **171**

24 O Teorema Espectral **173**

Capítulo 1

Introdução

Neste livro iremos estudar a matemática dos fenômenos lineares. Um fenômeno da natureza pode ser "entendido"como linear se a sua evolução ao longo do tempo é sempre a mesma em todas as escalas, como uma função cujo gráfico é uma linha reta. Essa ideia que está presente em inúmeros problemas do mundo real é encapsulada nos conceitos abstratos de espaços vetoriais e funções lineares entre espaços vetoriais e a partir daí sua aplicabilidade é quase ubíqua em física, engenharia, economia, computação, biologia e muito mais.

Iremos estudar basicamente dois objetos matemáticos: vetores e matrizes. Possivelmente você já ouviu falar deles no ensino médio, ou na faculdade. No entanto, nosso objetivo é mostrar como o entendimento profundo desses objetos pode revelar *padrões* em coisas tão diversas, desde o mercado de ações, a economia, ferramentas de buscas na internet... quase tudo!!!

Isso porque, este curso está orientado de forma a alinhar intuição, rigor matemático e aplicações. Desse modo, apresentamos os conceitos de forma rigorosa para em seguida darmos explicações intuitivas e inúmeras aplicações.

Assim sendo, vetores são vistos como pontos em espaços de dimensão arbitrária e matrizes como sendo transformações entre estes espaços. Esta linguagem geométrica é extremamente rica e permite uma melhor compreensão dos objetos em cena, o que traz consequências surpreendentes.

No restante desta introdução, daremos algumas ilustrações de como os conceitos que serão vistos ao longo do livro podem ser usados. Muitas destas aplicações serão retomadas e re-estudas no decorrer do curso.

Além disso, essa introdução também foi preparada com o intuito de enfatizar o poder das *demonstrações* para explicar, sintetizar e clarificar padrões e fenômenos complicados. Justamente porque o rigor matemático, um dos pilares desse livro, está aqui não por acaso!

1.1 O Teorema de Sylvester-Gallai

Vamos utilizar um exemplo muito simples para mostrar como uma demonstração elegante é capaz de explicar, sintetizar e clarificar um padrão complicado de se determinar.

Considere um conjunto com mais três de pontos no plano. Você pode desenhar num papel quatro, cinco ou seis pontos e irá notar que sempre acontece uma das coisas seguintes:

1. ou todos os pontos estão alinhados;
2. ou eles não estão alinhados, e é possível traçar uma reta que contenha **exatamente** dois dos pontos.

Ou seja: o pequeno experimento com alguns pontos desenhados aleatoriamente no plano revelou um padrão *potencial* se os pontos não estão alinhados então é possível traçar uma reta contendo dois, e não mais que dois, dos pontos.

No entanto, como verificar esse padrão num conjunto com três bilhões de pontos?

Teorema 1.1.1 (Sylvester-Gallai)**.** *Seja $\mathcal{C}$ um conjunto de n pontos no plano. Suponha que eles não estejam alinhados. Então existe uma reta que contém exatamente dois pontos do conjunto $\mathcal{C}$.*

Demonstração. Suponha por contradição que o teorema seja falso. Então, existe um conjunto de pontos $\mathcal{C}$ no plano tal que: os seus elementos não estão alinhados e toda reta que contém dois pontos de $\mathcal{C}$ contém na verdade pelo menos três pontos de $\mathcal{C}$.

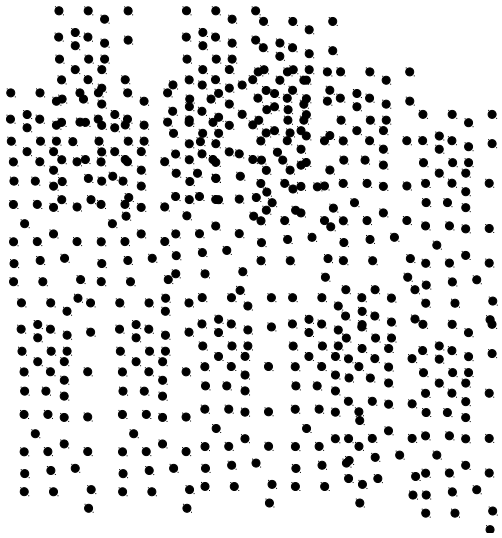

Figura 1.1: Encontre uma reta passa exatamente por dois pontos.

A abstração nos permite olhar para coisas que não são tão fáceis de enxergar à primeira vista. Assim, podemos ir além daquilo que os olhos nos mostram.

Pois bem. Cada dois pontos do conjunto $\mathcal{C}$ determinam uma reta no plano. Como os pontos de $\mathcal{C}$ não estão alinhados, para cada reta do plano com dois pontos de $\mathcal{C}$ existe um ponto fora dela. Assim, podemos considerar todos os pares ponto-reta (P, r) possíveis, com a reta r sendo gerada por dois pontos de $\mathcal{C}$ e P sendo um ponto de $\mathcal{C}$ fora da reta r.

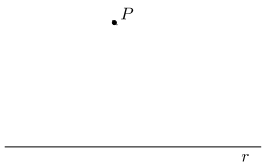

Figura 1.2: P é um ponto do conjunto $\mathcal{C}$ e r é uma reta gerada por dois pontos de $\mathcal{C}$.

Para cada par ponto-reta (P, r), podemos medir a distância entre o ponto e a reta. Como $\mathcal{C}$ tem finitos pontos, temos apenas um número finto de pares ponto-reta. Vamos olhar então para o par (P, r) com a menor distância entre o ponto P e a reta r.

Lembre que a distância de P a r é o comprimento do segmento PP' ortogonal a r, e P' é um ponto de r. Observe que o ponto P' divide a reta r em dois lados. Além disso, pela

hipótese de contradição, r contém pelo menos três pontos de $\mathcal{C}$. Assim um dos lados de P' contém ao menos dois pontos (um deles podendo ser o próprio P'.) Vamos supor que esse lado é o lado direito[1] (veja a Figura 1.3). Seja Q o ponto de $\mathcal{C}$ mais afastado de P' no lado direito.

Então, existe um ponto $B \in \mathcal{C}$, diferente de Q, mas que está entre P' e Q. Considere a reta ℓ gerada por P e Q. Marque em ℓ o ponto B' que dá a distância do ponto B à reta ℓ. Note que os triângulos $BB'Q$ e $PP'Q$ são semelhantes, sendo $BB'Q$ o menor deles. Isso prova que BB' é menor do que PP'. Mas isso prova justamente que o par ponto reta (B, ℓ)

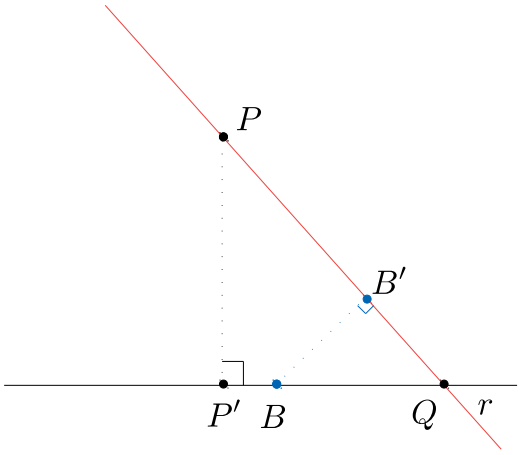

Figura 1.3: Encontrando um par ponto reta com a distância menor

possui uma distância menor do que o par (P, r). Mas o par (P, r) havia sido escolhido como aquele que minimizava essa distância. Isso dá uma contradição, e prova o teorema. □

Essa demonstração é linda! Mas além disso, ela *explica* porque tem que haver uma reta contendo exatamente dois dos pontos: se não houvesse sempre seria possível diminuir a distância reta-ponto dentro do conjunto $\mathcal{C}$, mas isso não pode ocorrer indefinidamente pois $\mathcal{C}$ é um conjunto finito.

[1]Se fosse do lado esquerdo o argumento seria idêntico. O leitor está convidado a reproduzir os desenhos nesse caso

1.2 *Netflix chalenge*

Uma das coisas que a maioria dos profissionais pretendem em suas carreiras é ganhar dinheiro? Já imaginou ganhar 1 milhão de dólares graças aos seus conhecimentos de Álgebra Linear? Pois, a alguns anos atrás a empresa americana *Netflix* lançou um desafio que oferecia essa soma como recompensa para o ganhador.

O desafio era melhorar o algorítimo de sugestão de filmes usado pela empresa. Funciona assim. Cada vez que um usuário assiste um filme ele dá uma nota de 1 a 5 estrelas para o filme. O algorítimo tenta então adivinhar a nota que o usuário dará pelo filme que ele (ainda) não assistiu!

Mais precisamente, cada usuário recebe sugestão de filmes para assistir. Quem decide o que sugerir para cada usuário é o algorítimo. Antes de lançar o desafio, menos de 8% dos filmes sugeridos eram de fato vistos. O desafio era aumentar para 10% essa proporção. Naturalmente, quando mais apuradas forem as sugestões mais pessoas irão fazer uma assinatura *Netflix*.

Mas então, como prever a nota que uma pessoa vai dar por um filme? É aí que entra a matemática. Imagine que colocamos as notas dos usuários em uma tabela (uma matrix) em que cada coluna representa um filme e cada linha um usuário.

Usuário/Filme	O senhor dos Anéis	Matrix	Jango
João	1	*	*
Livia	*	4	*

Ou seja, a matriz assim obtida contém pouca informação. Em 2009, uma equipe multinacional de engenheiros e cientistas da computação venceu o desafio ao utilisar um objeto geométrico associado a uma matriz: a decomposição em valores singulares. Essa decomposição mostra como uma matriz transforma esferas de grande dimensão em elipses. Usando esta informação matemática de modo crucial, a equipe conseguiu prever muito mais entradas não preenchidas que resultaram numa eficácia de 10% do algorítimo de sugestões da *Netflix*.

Neste curso, iremos aprender um pouco sobre a decomposição em valores singulares de uma matriz.

Capítulo 2

Espaços vetoriais

Espaços vetoriais são o palco onde matrizes atuam transformando vetores. É um conceito matemático abstrato que tenta capturar a essência do que deve ser uma "estrutura linear". Justamente por ser abstrato é um conceito flexível que aparece em inúmeras situações, e muitas vezes muito bem camuflado. Por isso um bom entendimento do conceito e de seus exemplos é fundamental para ter sucesso em aplicações.

2.1 Introdução

Definição 2.1.1. *Um* espaço vetorial *é um conjunto E, cujos elementos são chamados* vetores*, no qual estão definidas duas operações*

1. soma de vetores *que a cada par $u, v \in E$ associa um novo elemento de E, o qual denotamos por $u + v$.*

2. multiplicação de vetores por um escalar *que a cada elemento $v \in E$ e a cada número real $\lambda \in \mathbb{R}$ associa um novo elemento de E, o qual denotamos por λv.*

Estas operações devem ainda satisfazer os seguintes axiomas:

- ***comutatividade:*** $u + v = v + u$;

- ***associatividade:*** $(u+v)+w = u+(v+w)$*;*
- ***vetor nulo:*** *existe um vetor* $0 \in E$ *tal que* $0+u = u+0 = u$*, para todo* $u \in E$*;*
- ***simétrico aditivo:*** *para cada* $u \in E$ *existe um vetor* $-u \in E$ *que satisfaz* $u+(-u) = -u+u = 0$*;*
- ***distributividade:*** *se* $\alpha, \beta \in \mathbb{R}$ *e* $u \in E$ *então*

  $$(\alpha+\beta)u = \alpha u + \beta u$$

  e se, além disso, $v \in E$ *então*

  $$\alpha(u+v) = \alpha u + \alpha v;$$

- ***neutro multiplicativo:*** $1.u = u$*, para todo* $u \in E$*.*

Para exibir exemplos de espaços vetoriais precisamos exibir, em primeiro lugar, um conjunto E, e em segundo lugar devemos exibir uma maneira de somar os elementos desse conjunto e uma maneira de multiplica-los, ou "esticá-los", por números reais. É importante salientar que ao fazermos isso devemos sempre nos certificar que a soma de dois elementos de E é ainda um elemento de E, e analogamente para a multiplicação por um número real. Em outras palavras, ao declararmos como são as operações de soma e multiplicação por um número real devemos sempre nos certificarmos que o conjunto E é estável pelas operações que definimos.

Além disso, precisamos nos certificar que as operações que definimos em E satisfazem os axiomas de espaço vetorial.

2.2 Espaços Euclideanos

O exemplo mais importante de espaço vetorial nesse livro é dos espaços euclideanos. Os elementos desses espaços são listas ordenadas finitas de números. Podemos dizer a mesma coisa mais formalmente.

Definição 2.2.1. *O* espaço euclideano $\mathbb{R}^d$ *é conjuntos das d-uplas ordenadas de números reais:*

$$\mathbb{R}^d := \{(x_1, x_2, ..., x_d);\ x_1, x_2, ..., x_d \in \mathbb{R}\}.$$

Cada elemento $x = (x_1, x_2, ..., x_d) \in \mathbb{R}^d$ *é chamado um* vetor*, e os números* $x_1, ..., x_d$ *são chamados* coordenadas *do vetor* x*. O ponto* $\boldsymbol{0} = (0, 0, ..., 0)$ *é chamado a* origem*, ou o* vetor nulo*.*

Alguns vetores (em espaços euclideanos) têm importância especial (por razões que vão se tornar claras ao longo do livro). Por exemplo, os vetores *canônicos*

$$e_1 \stackrel{\text{def}}{=} (1, 0, ..., 0),\ e_2 \stackrel{\text{def}}{=} (0, 1, 0, ..., 0),\ ...\ e_d \stackrel{\text{def}}{=} (0, ..., 0, 1).$$

Ou seja, o vetor canônico e_j têm todas as entradas nulas, exceto a entrada j que é igual a 1. Um outro vetor importante é o *vetor de uns* $\mathbb{I} \stackrel{\text{def}}{=} (1, ..., 1)$, que tem todas as suas entradas iguais a 1. Nesse caso, observe que o tamanho do vetor de uns depende do contexto.

2.2.1 Exemplos de aplicações

Um vetor de tamanho d pode ser usado para representar d quantidades, ou d valores numéricos em alguma aplicação.

Exemplo 2.2.2 (Localização e deslocamento)**.** *Vetores de tamanho 2 e 3 podem ser usados para representar localização e deslocamento no plano e no espaço, respectivamente. Por exemplo, o vetor* $x = (2, -2, 1)$*, representado na figura 2.1, pode representar uma posição na superfície da terra em relação a alguma origem fixada e três eixos perpendiculares, mas também pode representar um deslocamento (basta, a partir de qualquer ponto, seguir na mesma direção, no mesmo sentido de* x*, um deslocamento que tenha o mesmo comprimento de* x*.)*[1]

Exemplo 2.2.3 (Cor)**.** *Um vetor de tamanho* 3 *pode representar a cor de um pixel numa imagem, no sistema RGB. Por exemplo, um pixel associado ao vetor* $(0, 1, 0)$ *é verde. Um pixel* $(1, 0.5, 0.5)$ *é um tom de rosa.*

[1] Vamos ver nos próximos capítulos como medir o comprimento de um vetor em geral.

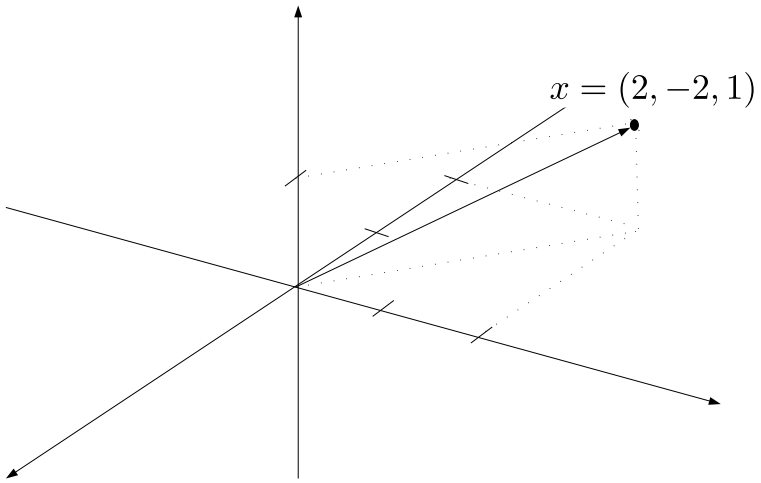

Figura 2.1: Na figura, está representado o vetor $x = (2, -2, 1)$, o qual é um elemento do espaço $\mathbb{R}^3$.

Exemplo 2.2.4 (Portfólio)**.** *Um vetor de $x = (x_1, ..., x_n)$ tamanho n pode representar um portfólio de investimento em n ativos diferentes, com o número x_i representando o número de ações do ativo i compradas pelo investidor. As entradas também podem ser frações (porcentagens) do total investido que são dedicadas a cada um dos ativos.*

Exemplo 2.2.5 (Aplicações médicas)**.** *Um vetor $x \in \mathbb{R}^d$ pode fornecer a quantidade de algum parâmetro de saúde ao longo de uma população de d indivíduos. Por exemplo, a pressão sanguínea de cada uma coleção de d pacientes, e nesse caso a entrada x_j fornece a pressão do paciente j.*

Exemplo 2.2.6 (Série temporal)**.** *Um vetor de tamanho n pode ser representar o valor de uma determinada quantidade ao longo de n intervalos temporais. Por exemplo, algum elemento $i \in \mathbb{R}^{50}$ pode fornecer taxa de inflação do Brasil nos últimos 50 anos.*

Exemplo 2.2.7 (Contador de palavras e histograma)**.** *Um vetor de tamanho n pode representar a quantidade de vezes que cada palavra de um dicionário de n palavras ocorre em um documento. Uma variação disso é o* histograma *de frequência de palavras: cada coordenada x_i fornece a frequência de aparição da i-ésima palavra no documento. Por exemplo, se $x \in \mathbb{R}^n$ é o histograma de um texto então $x_6 = 0.0005$ significa que a 6^a palavra do dicionário apareceu apenas 0.05% das vezes. Se x representasse a quantidade de palavras, e*

se o nosso texto tivesse 4000 palavras então x_6 seria igual $2 = 0.05\% \times 4000$. Contadores de palavras e histogramas de frequência são usadas em programas de computador que fazem análise automatizada de texto.

Os exemplos acima são apenas uma pequena amostra das potencialidades de aplicações. Motivados por isso, vamos estudar mais a fundo os espaços euclideanos.

2.2.2 Espaços euclideanos são exemplos de espaços vetoriais

Dados dois vetores x e y em um mesmo espaço euclideano podemos obter um novo vetor a partir deles, conforme a definição a seguir.

Definição 2.2.8. *Sejam $x = (x_1, x_2, ..., x_d) \in \mathbb{R}^d$ e $y = (y_1, y_2, ..., y_d) \in \mathbb{R}^d$. O vetor* soma *de x e y, denotado por $x + y$, é o vetor cujas coordenadas são $(x_1 + y_1, x_2 + y_2, ..., x_d + y_d)$. Em símbolos,*

$$x + y = (x_1 + y_1, x_2 + y_2, ..., x_d + y_d).$$

Portanto, operacionalmente somar dois vetores consiste simplesmente em somar-se as respectivas coordenadas.

Exemplo 2.2.9. *1. Sejam $x = (1, -2)$ e $y = (2, -1)$ em $\mathbb{R}^2$. Então $x + y = (3, -3)$*

2. Sejam $x = (0, 0, \sqrt{7}, 89, 14)$ e $y = (\pi, 1 + \sqrt{2}, 0, 0, 0)$ vetores em $\mathbb{R}^5$. Então,

$$x + y = (\pi, 1 + \sqrt{2}, \sqrt{7}, 89, 14).$$

A operação de soma de vetores num espaço euclideano $\mathbb{R}^d$ satisfaz os axiomas da definição de vetorial:

- **comutatividade** $x + y = (x_1 + y_1, ..., x_d + y_d) = (y_1 + x_1, ..., y_d + x_d)$, pois a soma de números é comutativa.
- **associatividade** A justificativa é análoga a anterior $x + (y + z) = (x + y) + z$ pois a soma de números é associativa
- **elemento neutro** $0 + x = x + 0 = (x_1 + 0, ..., x_d + 0) = (x_1, ..., x_d) = x$
- **simétrico aditivo** Analogamente, vemos que $-x + x = 0$

Exemplos de aplicações–

(a) **Deslocamentos** Se $a, b \in \mathbb{R}^2$ representam deslocamentos, então o vetor soma $a + b$ representa o deslocamento líquido obtido-se primeiro deslocando-se segundo o vetor a e em seguida deslocando-se de acordo com o vetor b **meter uma figura aqui**

(b) **Deslocamento entre dois pontos** Se $p, q \in \mathbb{R}^2$ representam localizações em uma mapa então o vetor $p - q$ representa o deslocamento de com início em q e término em p.

(c) **Contadores de palavras** Se $x, y \in \mathbb{R}^n$ representam as contagem de palavras de dois documentos (com todas as palavras pertencentes a um dicionário de n palavras) então $x + y$ é o contador de palavras do documento obtido juntando-se os dois. Já o vetor $x-y$ expressa a quantidade de vezes a mais que cada palavra ocorreu de um documento em relação a outro.

Definição 2.2.10. *Sejam $x \in \mathbb{R}^d$ um vetor e $\lambda \in \mathbb{R}$ um número real. O produto de λ por x é o vetor $\lambda x = (\lambda x_1, \lambda x_2, ..., \lambda x_3)$.*

Operacionalmente, mais uma vez, basta multiplicar coordenada a coordenada pelo número real em questão.

Exemplo 2.2.11. *1. Se $x = (1, 2, 1, -7) \in \mathbb{R}^4$ e $\lambda = 9 \in \mathbb{R}$ então $\lambda x = (9, 18, 9, -63)$.*

2. Se $x = (1, \sqrt{\pi}, \pi) \in \mathbb{R}^3$ e $\lambda = \sqrt{\pi}$ então $\lambda x = (\sqrt{\pi}, \pi, \pi\sqrt{\pi})$.

De forma análoga, podemos verificar que a operação de multiplicação por escalar satisfaz os demais axiomas de espaço vetorial.

Com as operações definidas, podemos verificar que $\mathbb{R}^d$ satisfaz os axiomas de espaço vetorial. Portanto, para cada $d \in \mathbb{N}$, o conjunto $\mathbb{R}^d$ é um espaço vetorial.

2.3 Outros exemplos de espaços vetoriais

Exemplo 2.3.1 (Solução de equações)**.** *Muitos problemas da natureza podem ser modelados como equações diferenciais, equações que envolvem a derivada uma função. O conjunto de*

soluções de algumas equações diferenciais é um espaço vetorial, como veremos em capítulos posteriores.

Exemplo 2.3.2 (Espaço de Funções)**.** *Considere $F(\mathbb{R},\mathbb{R})$ o conjunto de todas as funções $f:\mathbb{R}\to\mathbb{R}$. Evidentemente podemos somar funções e multiplicar funções por números reais: a cada par $f,g\in C^0(\mathbb{R},\mathbb{R})$, a função soma $f+g:\mathbb{R}\to\mathbb{R}$ é definida por*

$$(f+g)(x)=f(x)+g(x).$$

Analogamente, dados $\lambda\in\mathbb{R}$ e $f\in C^0(\mathbb{R},\mathbb{R})$, a função $\lambda f:\mathbb{R}\to\mathbb{R}$ é definida por

$$(\lambda f)(x)=\lambda f(x).$$

Além disso, sabemos do cálculo que a soma de duas funções contínuas e produto de funções contínuas sempre fornecem funções contínuas. Portanto, $C^0(\mathbb{R},\mathbb{R})$ é estável pelas operações que definimos.

Iremos verificar a seguir que o conjunto $C^0(\mathbb{R},\mathbb{R})$ com as operações definidas acima satisfaz os axiomas de um espaço vetorial. Em muitas dessas demonstrações vamos provar que duas funções f e g são iguais. Lembre que para isso é necessário e suficiente mostrar que os valores assumidos por f e g são sempre os mesmos. Ou seja, para todo $x\in\mathbb{R}$, vale que $f(x)=g(x)$.

1. ***comutatividade:*** *para mostrar que $f+g=g+f$, baste verificar que $f(x)+g(x)=g(x)+f(x)$, para todo $x\in\mathbb{R}$. Mas isso é verdade pois a adição de números reais é comutativa.*

2. ***associatividade:*** *para mostrar que $(f+g)+h=f+(g+h)$ precisamos verificar que $(f(x)+g(x))+h(x)=f(x)+(g(x)+h(x))$, para todo $x\in\mathbb{R}$, o que é verdade pois a adição de números reais é associativa.*

3. ***vetor nulo:*** *a função constante igual a zero, $0(x)=0$ é o vetor nulo em $C^0(\mathbb{R},\mathbb{R})$. De fato, $0+f=f+0$ pois*

$$f(x)+0(x)=f(x)+0=f(x)=0+f(x)=0(x)+f(x),$$

para todo $x\in\mathbb{R}$.

4. ***simétrico aditivo:*** *o simétrico de $f \in C^0(\mathbb{R},\mathbb{R})$ é a função $-f$, definida por $(-f)(x) = -f(x)$. Vale que $f + (-f) = -f + f = 0$ pois $f(x) + (-f(x)) = -f(x) + f(x) = 0$.*

5. ***distributividade:*** *sejam $\alpha, \beta \in \mathbb{R}$ e $f \in C^0(\mathbb{R},\mathbb{R})$. Vamos verificar que*

$$(\alpha + \beta)f = \alpha f + \beta f.$$

Novamente, temos uma função do lado esquerdo e uma função do lado direito e queremos checar que são iguais. Basta ver que, para cada $x \in \mathbb{R}$, as funções dão o mesmo valor. De fato,

$$(\alpha + \beta)(f)(x) = (\alpha + \beta)f(x) = \alpha f(x) + \beta f(x) = (\alpha f + \beta f)(x).$$

6. ***neutro multiplicativo:*** *Queremos verificar que $1.f = f$. Mas $1.f(x) = f(x)$, para todo $x \in \mathbb{R}$, então segue.*

Enquanto vetores de um espaço euclideano podem modelar séries temporais *finitas e discretas* o exemplo acima modela séries temporais *infinitas e contínuas*.

Exemplo 2.3.3 (O espaço dos polinômios)**.** *Temos subconjuntos especiais de $C(\mathbb{R},\mathbb{R})$:*

$$\mathcal{P}(\mathbb{R}) = \{p \in C(\mathbb{R},\mathbb{R}); \exists\, a_0, ..., a_n \in \mathbb{R}; p(x) = a_0 + a_1x + a_2x^2 + ... + a_nx^n,\ \forall\, x \in \mathbb{R}\}.$$

e

$$\mathcal{P}_d(\mathbb{R}) = \{p \in \mathcal{P}(\mathbb{R}); n \leq d\}.$$

Em palavras: um polinômio de grau n é uma função $p : \mathbb{R} \to \mathbb{R}$ da forma $p(x) = a_0 + a_1x + a_2x^2 + ... + a_nx^n$, onde $a_0, ..., a_n$ são números reais (parâmetros do polinômio). Ambos os conjuntos são espaços vetoriais. Deixamos a verificação desse fato a cargo do leitor. No entanto, essa verificação será muito mais fácil quando apresentarmos o conceito de subespaços: $\mathcal{P}(\mathbb{R})$ e $\mathcal{P}_d(\mathbb{R})$ são subespaços de $C^0(\mathbb{R},\mathbb{R})$.

De fato, veja que a soma de dois polinômios é ainda um polinômio, e que a multiplicação por número real de um polinômio dá um outro polinômio. O conjunto dos polinômios é estável pelas operações do espaço $C^0(\mathbb{R},\mathbb{R})$.

2.4 Combinações lineares

Num espaço vetorial podemos somar vetores e "esticá-los", multiplicando-os por números reais. Além disso, o axioma de associatividade diz que podemos somar três vetores, quatro, etc... e mais geralmente qualquer número n de vetores. Assim, dados n vetores $v_1, ..., v_n \in E$, podemos esticar cada um deles e depois somar tudo.

Definição 2.4.1. *Dados n vetores $v_1, ..., v_n$ e n números reais $\lambda_1, ..., \lambda_n$, o vetor*

$$v = \lambda_1 v_1 + ... + \lambda_n v_n$$

é chamado uma combinação linear *dos vetores $v_1, ..., v_n$.*

Observação 2.4.2. *Por questões de estética vamos utilizar neste livro a notação de somatório*

$$\sum_{i=1}^{n} \lambda_i = \lambda_1 + \lambda_2 + ... + \lambda_n,$$

que torna muito concisa e elegante expressões como combinações lineares, que serão recorrentes ao longo do livro.

Exemplo 2.4.3.

1. *Sejam $e_1 = (1, 0)$ e $e_2 = (0, 1)$ vetores em $\mathbb{R}^2$. Então $v = (3, 7)$ é combinação linear de e_1 e e_2. Com efeito,*

$$3e_1 + 7e_2 = (3, 0) + (0, 7) = (3, 7).$$

2. *Mais geralmente, todo vetor $v = (x_1, ..., x_d) \in \mathbb{R}^d$ é combinação linear dos vetores $e_1 = (1, 0, 0, ..., 0), e_2 = (0, 1, 0, 0, ..., 0), ..., e_d = (0, 0, ..., 1)$, pois*

$$\sum_{j=1}^{n} x_j e_j = (x_1, 0, ...0) + ... + (0, 0, ..., x_d) = (x_1, ..., x_d).$$

3. *Em $\mathbb{R}^3$, $v = (3 + 5\pi, 6, 31)$ é combinação linear dos vetores $v_1 = (1, 2, -4)$ e $v_2 = (\pi, 0, 7)$ pois*

$$3v_1 + 5v_2 = (3, 6, -4) + (5\pi, 0, 35) = (3 + 5\pi, 6, 31).$$

4. *No espaço $C^0(\mathbb{R}, \mathbb{R})$ das funções contínuas, sejam $f(t) = t^2$ e $g(t) = \sin(t)$. Então*

$$h(t) = 3t^2 + 9\sin(t)$$

é combinação linear de f e g.

2.4.1 Algumas combinações lineares especiais

Seja E um espaço vetorial e sejam $v_1, ..., v_n$ elementos de E. Considere uma combinação linear $x = \lambda_1 v_1 + ... + \lambda_n v_n$.

Soma Quando $\lambda_1 = \lambda_2 = ... = \lambda_n$, temos

$$x = v_1 + ... + v_n,$$

a *soma* dos vetores $v_1, ..., v_n$.

Média (ou centroide) Quando $\lambda_1 = \lambda_2 = ... = \lambda_n = 1/n$, dizemos que

$$x = \frac{1}{n} v_1 + ... + \frac{1}{n} v_n$$

é a média dos vetores. A média as vezes é chamada de *centroide* do conjunto finito $v_1, ..., v_n$, o que é uma acepção mais geométrica, relacionada ao fato de que a média do conjunto finito $v_1, ..., v_n$ é o ponto do espaço vetorial que está simultaneamente mais perto de todos os elementos do conjunto. Vamos enunciar e demonstrar rigorosamente esse fato no Capítulo 5, onde ele será fundamental para o entendimento de um algoritmo muito usado em aplicações em Ciência de Dados.

Combinação convexa Se os coeficientes somam 1, isto é, $\sum_{\ell=1}^{n} \lambda_\ell = 1$, então x é chamado de *combinação afim* dos vetores $v_1, ..., v_n$. Se, além disso, cada $\lambda_\ell \geq 0$ então x é uma *combinação convexa* ou *média ponderada* dos vetores. Observe que nesse caso, os coeficientes necessariamente números entre 0 e 1 e representam pesos, ou porcentagens que somam 100%.

2.5 Julia

Nessa versão beta do livro eu não vou poder fornecer um manual de introdução a Julia. No entanto vou colocar algumas dicas do que fazer (e um pouco de como fazer) em Julia para complementar e **abrilhantar** o seu aprendizado de Álgebra Linear.

2.5.1 Instalação

Acredite: eu sou uma topeira com computadores e eu consegui instalar Julia. Então você consegue. Toma um tempo, guarde uma tarde agradável de domingo quando você tenha disponibilidade e paciência para isso. Vou falar o que eu fiz. Se você já é um entendedor de linguagem de programações, faça como achar melhor.

A primeira coisa é baixar a distribuição anaconda `https://www.anaconda.com/download/`. Depois você baixa o Julia em si: `https://julialang.org/downloads/`.

Feito isso você pode escrever seus programas direto no prompt de comando do Julia, ou usando um Jupyter notebook que você abre no anaconda.

2.5.2 Usando Julia

Você pode usar o Julia, no mínimo, como uma super calculadora para fazer contas com vetores grandes, por exemplo.

Você pode calcular a soma de dois vetores em $\mathbb{R}^7$:

x=[3.78,7.98,8.67,778.67,8.777,666,4.324]
y=[1,2,3,4.5,6.7,8.9,666]
x+y

Se estiver usando o Jupyter, para "rodar"o seu programa você precisa sempre apertar shit+enter no teclado. Ao fazer isso você deve obter como resposta

7-element ArrayFloat64,1:
4.779999999999999
9.98
11.67
783.17
15.477
674.9

670.324

Se você digita 5x+89y e dá shit+enter ele te devolve o resultado dessa combinação linear:

7-element ArrayFloat64,1:

107.9

217.9

310.35

4293.85

640.1850000000001

4122.1

59295.62

Uma coisa muito legal que Julia pode fazer para ajudar você "ver"um vetor em $\mathbb{R}^7$, $\mathbb{R}^{50}$ por exemplo (ou qualquer em qualquer espaço euclideano $\mathbb{R}^d$) é interpretar o vetor como o gráfico de uma função $x : [1, d] \to \mathbb{R}$ cujo domínio é o conjunto $\{1, ..., d\}$ e o contradomínio é o conjunto dos números reais. Você pode visualizar o vetor $x \in \mathbb{R}^7$ que eu escrevi acima usando a ferramenta Plots do Julia.

No Jupyter, primeiro você vai adicionar o pacote Plots, com a linha de comando:

```
import Pkg
Pkg.add("Plots")
```

Espera um pouco até compilar (você vai ver que some o asterisco em cima da região onde você digita o comando). Depois você tem que avisar que vai querer usar o Plots nesse seu notebook dizendo

```
using Plots
```

Pronto. Agora é só plotar seus vetores

```
plot(x, marker = :circle, legend = false, grid = false)
```

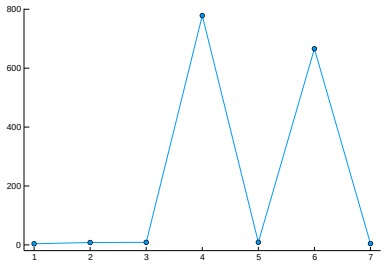

Figura 2.2: Para salvar a figura você tem que adicionar o comando savefig("nome.pdf"). A figura fica salva no diretório do anaconda.

O que retorna a figura2.2 a seguir

Olhar um vetor como um gráfico já nos permite contemplar a variação brutal de magnitude entre uma coordenada a outra. Você vê automaticamente qual a maior entrada, qual a menor entrada, qual a médias das entradas. Adiante no livro vamos ver como essas informações são relvantes nas aplicações. Outras coisas divertidas que você pode fazer gerar e plotar um vetor aleatório $z \in \mathbb{R}^{50}$, gerar e plotar um vetor canônico qualquer de $\mathbb{R}^{57}$. Eu escolhi o e_8.

j=8; d=57;

ej=zeros(d) ✓Aqui você declara que o ej tem todas as entradas nulas

ej[j]=1 ✓aqui você muda de ideia e manda a j-ésima entrada do ej ser igual a 1

plot(ej,marker = :circle, legend = false, grid = false) ✓Aqui você plota.

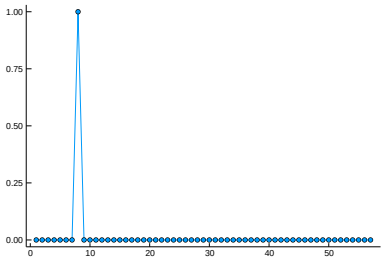

Figura 2.3: Oitavo vetor canônico de $\mathbb{R}^{57}$ visto como um gráfico

Capítulo 3

Produto interno

Neste capítulo vamos introduzir um pedaço adicional de estrutura que se pode colocar num espaço vetorial, o *produto interno*. Com o produto interno podemos fazer geometria em espaços vetoriais: medir comprimento de vetores, ângulo entre vetores, fazer projeções ortogonais e muitas outras coisas. Vamos mostrar também, já nesse capítulo e nos subsequentes, como essa geometria de vetores de tamanho 10 milhões têm impressionantes aplicações.

3.1 Introdução

Definição 3.1.1. *Seja E um espaço vetorial. Um* produto interno *em E é uma função*

$$\begin{aligned} \langle , \rangle : E \times E &\to \mathbb{R} \\ (u, v) &\mapsto \langle x, y \rangle \in \mathbb{R} \end{aligned}$$

que a cada par de vetores $u, v \in E$ associa o número real $\langle x, y \rangle$, de modo a satisfazer as seguintes propriedades:

- ***simetria*** $\langle x, y \rangle = \langle y, x \rangle$*;*
- ***linearidade*** $\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle$*, para todos $\alpha, \beta \in \mathbb{R}$.*
- ***positividade definida*** $\langle x, x \rangle \geq 0$ *e* $\langle x, x \rangle = 0$ *se e somente se $x = 0$, o vetor nulo de E.*

Então, para exibir um exemplo de um produto interno na verdade primeiro precisamos exibir um espaço vetorial E e em seguida devemos definir um função definida no produto cartesiano do espaço por ele mesmo $E \times E$ tomando valores em $\mathbb{R}$ e que seja uma função simétrica, linear e positiva definida. Como o exemplo mais importante de espaço vetorial são os espaços euclideanos, é ali que vamos exibir o produto interno mais importante desse livro.

3.2 Produto interno euclideano

Nesta seção vamos considerar apenas os espaços vetoriais euclideanos $\mathbb{R}^d$. Então vamos fixar um número natural d qualquer e definir um produto interno em $\mathbb{R}^d$.

Definição 3.2.1. *Sejam $x, y \in \mathbb{R}^d$. O* produto interno euclideano *de x e y é o número*

$$\langle x, y\rangle \stackrel{\text{def}}{=} \sum_{\ell=1}^{d} x_\ell y_\ell = x_1 y_1 + ... + x_d y_d.$$

Ou seja, você multiplica coordenada a coordenada e depois soma tudo. Como a ordem dos fatores não altera um produto numérico, é claro que essa função é simétrica. Além disso, pela distributividade da adicção em relação a multiplicação podemos verificar que essa função é linear. Com efeito, se $x, y, z \in \mathbb{R}^d$ e $\alpha, \beta \in \mathbb{R}$ então

$$\langle \alpha x + \beta y, z\rangle = \sum_{\ell=1}^{d} (\alpha x_\ell + \beta y_\ell) z_\ell = \alpha \sum_{\ell=1}^{d} x_\ell z_\ell + \beta \sum_{\ell=1}^{d} y_\ell z_\ell = \alpha\langle x, z\rangle + \beta\langle y, z\rangle.$$

Quase sempre nesse livro esse vai ser o único produto interno que vamos usar num espaço euclideano. Então, vamos fixar que se $x, y \in \mathbb{R}^d$ então o símbolo $\langle x, y\rangle$ sempre vai denotar o produto interno euclideano como definido acima. Outros produtos internos vão aparecer aqui e acolá eventualmente, mas quando for a hora vamos ser explícitos.

3.2.1 Alguns casos especiais

Seja $x \in \mathbb{R}^d$ um vetor arbitrário.

- **vetores canônicos** seja e_j um vetor canônico. Então $\langle e_j, x\rangle = x_j$; ou seja, o produto interno com o j-ésimo vetor canônico retorna a j-ésima coordenada de x.

- **vetor de uns** seja $\mathbb{I} = (1, ..., 1) \in \mathbb{R}^d$. Então, o produto interno $\langle x, \mathbb{I}\rangle = \sum_{\ell=1}^{d} x_\ell$ fornece a soma das coordenadas de x.

- **média** seja $m = (1/d, ..., 1/d)$. Então $\langle x, m\rangle = \frac{1}{d}\sum_{\ell=1}^{d} x_\ell$ é a média das coordenadas de x.

- **soma dos quadrados** O produto interno de x por ele mesmo $\langle x, x\rangle = \sum_{\ell=1}^{d} x_\ell^2$ fornece a soma dos quadrados das entradas de x. Em particular, isso prova que o produto interno euclideano é uma função positiva definida.

- **soma seletiva** seja $y = (1, 0, 0, ..., 0, 1) \in \mathbb{R}^d$. Então $\langle x, y\rangle = x_1 + x_d$. Mais geralmente, se y tem algumas entradas iguais a 1 e as demais nulas, então o produto interno $\langle x, y\rangle$ fornece a soma de apenas algumas entradas de x (aquelas correspondentes às entradas não-nulas de y).

3.2.2 Aplicações

Vamos a seguir ver alguns modelos de situações onde o produto interno euclideano pode ser usado para obter conclusões interessantes.

Exemplo 3.2.2 (Co-ocorrência)**.** *Podemos usar vetores de 0's e 1's para representar ocorrência de um determinado evento ao longo de d medições. Por exemplo, imagine um conjunto finito com 7 elementos, $X = \{x_1, x_2, ..., x_7\}$, que pode representar um conjunto de pacientes num hospital, apenas para fixar ideias. Suponha que tenhamos dois subconjuntos A, B de X, representando por exemplo pacientes com câncer no pulmão e pacientes não-fumantes, conforme a figura abaixo Nesse caso, podemos formar vetores*[1] *que representem os subconjuntos: basta ascender a entrada j (colocando ela igual a 1) caso o elemento x_j pertença ao subconjunto, e apagar ela (colocando ela igual a 0 caso contrário). Assim, temos que os vetores*

$$a = (0, 1, 1, 1, 1, 1, 1) \in \mathbb{R}^7 \ \ e \ \ b = (1, 0, 1, 0, 1, 0, 0) \in \mathbb{R}^7$$

[1] Que são muito fáceis de se colocar no computador, e portanto de fazer contas!!!

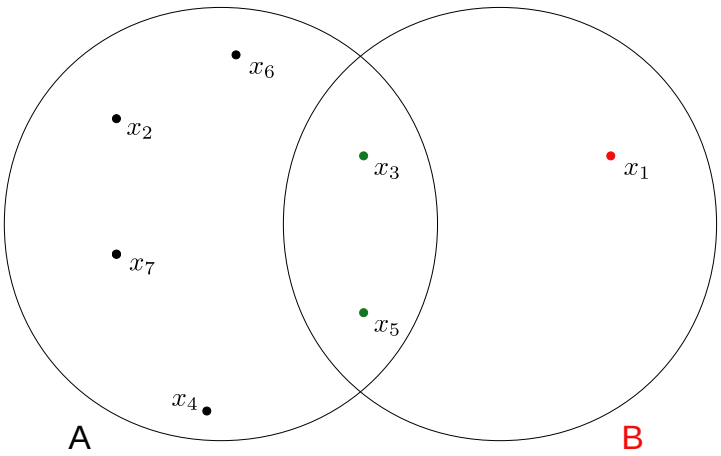

Figura 3.1:

representam, respectivamente os subconjuntos A e B. O produto interno fornece então a quantidade de elementos na interseção $A \cap B$:

$$\langle a, b \rangle = 2.$$

Em particular, podemos usar isso para contar quantos pacientes não-fumantes têm câncer de pulmão. O ponto importante desse exemplo é justamente a flexibilidade que advém da abstração. A quantidade de elementos do conjunto X pode ser arbitrariamente grande, e com uma linguagem de programação é muito fácil converter prontuários digitalizados em vetores de 0's e 1's. Ou seja, em questão de segundos podemos usar essa técnica para saber co-ocorrência de doenças e hábitos de vida, apenas para ilustrar uma possível aplicação.

Exemplo 3.2.3 (Microeconomia)**.** *Se $p \in \mathbb{R}^n$ é o* vetor de preços *das commodities de uma determinada economia (ou seja p_j é o preço da commoditie j), e $q \in \mathbb{R}^n$ é a* cesta de consumo *(ou seja q_j é quanto se compra da commoditie j) de um determinado consumidor então o produto interno $\langle p, q \rangle$ é o consumo de renda dessa cesta.*

Exemplo 3.2.4 (Probabilidade)**.** *Se $p \in\in \mathbb{R}^d$ é um vetor de probabilidades, ou seja $0 \leq p_j \leq 1$ para todo $j = 1, ..., d$ e $\sum_{j=1}^{d} p_j = 1$, e se $x \in \mathbb{R}^d$ é um vetor qualquer (uma variável aleatória na linguagem dos probabilistas) então o produto interno $\langle p, x \rangle$ é o valor esperado de x.*

Exemplo 3.2.5 (Evaluação de polinômios)**.** *Um polinômio* $p(t) = a_0 + a_1 t + a_2 t^2 + ... + a_d t^d$ *pode ser re-interpretado como um produto interno: basta considerar os vetores*

$$a = (a_0, ..., a_d) \in \mathbb{R}^{d+1} \quad e \quad T = (1, t, ..., t^d) \in \mathbb{R}^{d+1}.$$

Então

$$p(t) = \langle a, T \rangle.$$

Exemplo 3.2.6 (Análise automática de sentimentos num texto)**.** *Considere um dicionário com n palavras e um documento de texto com todas as palavras contidas nesse dicionário. Suponha que* $x \in \mathbb{R}^n$ *seja o vetor cuja entrada j forneça a quantidade de aparições da j-ésima palavra no documento. Usando uma linguagem de programação (Julia, Phyton, etc...) é muito fácil obter esse vetor. Suponha agora que façamos duas listas de palavras, uma contendo as palavras "negativas"do dicionário (mau, ruim, horrível, fodido, merda, chato, etc...) e outra com as palavras "positivas"(bom, legal, excelente, sensacional,excitante, etc...). Todas as outras palavras podem ser colocadas como neutras. Agora considere o vetor* $w \in \mathbb{R}^n$ *cujas entradas são dadas por*

$$w_j = \begin{cases} +1, \text{ se a j-ésima palavra é positiva,} \\ 0, \text{ se a j-ésima palavra é neutra,} \\ -1, \text{ se a j-ésima palavra é negativa} \end{cases}.$$

Então, o produto interno $\langle x, w \rangle$ *fornece uma medida do sentimento do documento: quanto mais positivo o produto interno vemos que o documento possui muito mais palavras positivas do que negativas; quanto mais negativo for, vemos que o documento possui muito mais palavras negativas do que positivas.*

3.3 Outros exemplos de produto interno

Vamos agora ver outras formar de se colocar a estrutura de produto interno num espaço vetorial.

3.3.1 Produto interno euclideano com pesos

Assim como demonstramos que o produto interno euclideano satisfaz os axiomas de simetria, linearidade e positividade definida, podemos verificar (deixo isso a cargo do leitor) que a função P seguir define um produto interno em $\mathbb{R}^d$: dados números reais não-negativos $p_1, ..., p_d$, considere a função $P : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ definida por

$$P(x, y) = \sum_{j=1}^{d} p_j x_j y_j, \ \ \forall \ x, y \in \mathbb{R}^d.$$

Observe que o produto interno euclideano é um caso particular dessa função quando $p_1 = p_2 = ... = p_d = 1$

3.3.2 Integral como produto interno

Esse exemplo é apenas para abrir sua mente a horizontes mais amplos. Vou assumir que você já viu alguma coisa de Cálculo 1. Se não viu, dê uma pesquisada na Internet para aprender o que os nomes significam. Considere o espaço vetorial das funções contínuas $C^0([0, 1], \mathbb{R})$. Ele é praticamente a mesma coisa do espaço das funções que apresentamos no primeiro capítulo, com a exceção de que as funções contínuas não podem ser arbitrariamente loucas. Em particular, todas elas são integráveis. Com, isso a integral define um produto interno nesse espaço:

$$I(f, g) \stackrel{\text{def}}{=} \int_0^1 f(t)g(t)dt.$$

Esse produto interno é muito importante em engenharia, por exemplo, na análise de frequência de sinais de telecomunicações.

Capítulo 4

Comprimento e ângulo

Agora vamos ver como usar o produto interno para fazer geometria com vetores. Como isso aqui é apenas uma introdução à Álgebra Linear, não iremos adentrar nas profundezas da geometria dos espaços vetoriais e por isso vamos nos ater apenas às ideias geométricas básicas de comprimento e ângulo. Com apenas isso já dá para fazer muita coisa maravilhosa. Por outro lado, quero apresentar a você, caro leitor, uma faceta sofisticada da matemática contemporânea, que na falta de um nome melhor, vou chamar de relativização dos conceitos clássicos.

Há muito tempo os matemáticos perceberam que flexibilizar a maneira como medimos coisas, como comprimento e ângulo, tem muitas aplicações não-triviais. Por essa razão, vamos adotar nesse curso uma abordagem axiomática a esse estudo geométrico.

4.1 Norma

A axiomatização da ideia de comprimento se faz através do conceito de norma.

Definição 4.1.1. *Seja E um espaço vetorial. Uma* norma *em E é uma função*

$$\begin{aligned} \|.\| : E &\to [0, +\infty) \\ v \in E &\mapsto \|v\| \in [0, +\infty), \end{aligned}$$

satisfazendo às três condições a seguir

- ***Positividade definida*** $\|v\| = 0 \iff v = 0$
- ***Homogeneidade*** $\|\lambda v\| = |\lambda|\|v\|, \ \forall\, \lambda \in \mathbb{R}$
- ***Desigualdade triangular*** $\|u + v\| \leq \|u\| + \|v\|, \ \forall\, u, v \in E.$

Ou seja, essa definição diz o seguinte: olha, seja lá como for que você queira medir o comprimento dos seus vetores, algumas coisas mínimas você vai ter que respeitar. Primeiro, a sua "maneira de medir comprimento"tem que ser uma coisa que pega um vetor e devolve um número não-negativo[1], o comprimento do vetor. Além disso todo vetor que não for o vetor nulo tem que ter comprimento estritamente positivo (e o vetor nulo tem que ter comprimento zero); se você esticar um vetor por um fator numérico, o comprimento fica esticado pelo mesmo fator, porém com a ressalva de que se você "inverter"o vetor (multiplicando ele por um número negativo) isso não pode mudar o comprimento[2] Por último, sua maneira de medir comprimento tem que se sujeitar a máxima da sabedoria popular: uma reta é sempre o caminho mais curto, em outras palavras a soma de dois dos lados de um triângulo é sempre maior ou igual do que o outro lado.

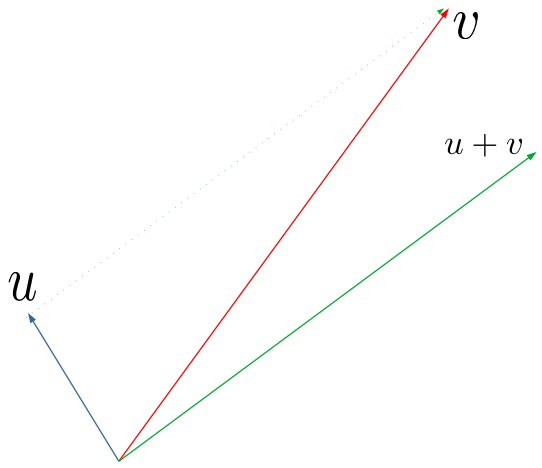

Figura 4.1: Desigualdade triangular: o comprimento de u somado ao comprimento de v não excede o comprimento da soma $u + v$.

[1] No jargão matemática, isso é uma função, cujo domínio é um espaço vetorial E e cujo contradomínio é o intervalo $[0, +\infty)$ dos números não-negativos

[2] Por isso aparece $|\lambda|$ na hipótese de homogeneidade.

Observação 4.1.2. ***Um comentário sobre notação:*** *o símbolo* $\|.\|$ *serve apenas para indicar o nome da função, e é um costume antigo usá-lo para denotar normas. Isso porque, como vamos ver já já, o conceito de norma generaliza o* módulo *de um número real, que é denotado por* $|x|$.

Vamos ver então exemplos de normas.

4.1.1 A norma do máximo

Vamos considerar um espaço euclideano $\mathbb{R}^d$ e definir a função

$$\|x\|_\infty \stackrel{\text{def}}{=} \max\{|x_j|; j = 1, ..., d\},$$

que calcula o maior valor absoluto das entradas do vetor x. Vamos verificar que essa função é uma norma:

- **Positividade definida:** é claro que a maior entrada de x é zero se, e somente se, $x = (0, ..., 0)$.
- **Homogeneidade** Dado $\lambda \in \mathbb{R}$, sabemos que $\lambda x = (\lambda x_1, ..., \lambda x_d)$, logo o maior valor absoluto dentre as entradas de λx é $|\lambda|\times$(maior valor absoluto dentre as entradas de x). Ou seja, $\|\lambda x\|_\infty = |\lambda|\|x\|_\infty$.
- **Desigualdade triangular** A desigualdade triangular é um pouco mais interessante. Olha só: sejam $x, y \in \mathbb{R}^d$. Sejam x_i e y_k, respectivamente as maiores entradas de x e de y em valor absoluto. Ou seja, $\|x\|_\infty = x_i$ e $\|y\|_\infty = y_k$. Por outro lado, as entradas de $x + y$ são os números $x_j + y_j$, com $j \in \{1, ..., d\}$. Como o módulo da soma de dois números reais é sempre menor do que ou igual a soma dos módulos desses números temos que

  $$|x_j + y_j| \leq |x_j| + |y_j|, \ \forall \ j \in \{1, ..., d\}.$$

  Por outro lado, $|x_j| \leq |x_i|$ e $|y_j| \leq |y_k|$, para todo j, uma vez que x_i e y_k são as maiores entradas. Juntando as desigualdades que temos chegamos a

  $$|x_j + y_j| \leq |x_i| + |y_k| = \|x\|_\infty + \|y\|_\infty,$$

  como queria demonstrar.

Exemplo 4.1.3. *Em* $\mathbb{R}^3$ *considere* $x = (1, -2, -7)$ *e* $y = (-7, 2, 6)$. *Então* $x+y = (6, 0, -1)$. *Logo,* $\|x\|_\infty = 7$, $\|y\|_\infty = 7$ *e* $\|x + y\|_\infty = 6$. *Como* $6 < 14$, *vemos a desigualdade triangular funcionar nesse exemplo! Agora, se* $x = (1, 2, 7)$ *e* $y = (7, 2, 6)$ *então* $x + y = (8, 4, 13)$. *Como* $13 < 14$ *vemos de novo se produzir diante dos nossos olhos a desigualdade* $\|x+y\|_\infty \leq \|x\|_\infty + \|y\|_\infty$.

4.1.2 A norma da soma (ou a distância do taxista)

Ainda num espaço euclideano $\mathbb{R}^d$ qualquer, temos uma outra função que serve como maneira de medir comprimento:

$$\|x\|_1 = \sum_{j=1}^{d} |x_j|.$$

Ou seja, essa função pega um vetor $x \in \mathbb{R}^d$ e retorna a soma dos valores absolutos das entradas de x. A verificação de que essa função é uma norma fica a cargo do leitor.

4.2 Normas advindas de um produto interno

O tipo mais importante de norma desse livro são aquelas que são definidas a partir de um produto interno. Considere E um espaço vetorial qualquer, munido de um produto interno $\langle , \rangle$. Considere a função

$$\|x\| = \sqrt{\langle x, x \rangle}, \ \forall \ x \in E.$$

Observe que agora não estamos mais fornecendo um exemplo explícito, como quando falamos dos espaços euclideanos. Estamos falando de um exemplo conceitual, de como a estrutura de produto interno permite definir uma nova estrutura capaz de medir comprimentos. Num certo sentido, como vamos ver logo a seguir, essa é a maneira "correta"no sentido clássico, porque ela é inspirada pelo teorema de Pitágoras. Bom, a demonstração de que essa função satisfaz os requisitos mínimos para ser considerada uma norma exige um esforço a mais de elevação intelectual, ao qual nos dedicamos a seguir.

4.2.1 Cauchy-Schwartz

É impressionante a constatação de que quando abstraímos as ideias e as transformamos em conceitos que capturam somente o essencial daquele fenômeno de interesse, tudo fica muito mais elegante límpido.

De fato, observe que a norma oriunda do produto interno é positiva definida porque o produto interno é também uma função positiva definida. A homogeneidade segue da linearidade do produto interno:

$$\|\lambda x\| = \sqrt{\langle \lambda x, \lambda x\rangle} = \sqrt{\lambda^2 \langle x, x\rangle} = |\lambda|\sqrt{\langle x, x\rangle} = |\lambda|\|x\|,$$

pois a raiz quadrada do quadrado de um número é sempre igual ao seu valor absoluto.

Para estabelecer a desigualdade triangular, vamos começar provando uma desigualdade clássica e muito importante que relaciona a norma e o produto interno[3]

Proposição 4.2.1 (Cauchy-Schwartz)**.** *Seja E um espaço vetorial munido de um produto interno $\langle .,. \rangle$ e seja $\|.\|$ a norma oriunda desse produto interno, isto é, $\|x\| = \sqrt{\langle x, x\rangle}$ para todo vetor $x \in E$. Então, quaisqer que sejam $x, y \in E$ vale que*

$$|\langle x, y\rangle| \leq \|x\|\|y\|.$$

Além disso, a igualdade acontece se, e somente se, x e y são colineares*, ou seja, se existe um número real $\lambda \in \mathbb{R}$ tal que $x = \lambda y$.*

Demonstração. Sejam $x, y \in E$ dois vetores quaisquer. Note que se um deles for o vetor nulo a desigualdade é óbvia (e ainda é uma igualdade nesse caso). Assim sendo, podemos assumir que nem x nem y são iguais ao vetor nulo de E.

Pela simplicidade da escrita de fórmulas e desigualdades, vamos denotar, ao longo dessa demonstração, $\xi = \|x\|$ e $\vartheta = \|y\|$, e podemos assumir que ξ e ϑ são números reais positivos, já que nem x nem y são nulos. A positividade definida do produto interno nos diz que

$$\|\xi y - \vartheta x\|^2 = \langle \xi y - \vartheta x, \xi y - \vartheta x\rangle \geq 0.$$

[3]Quando a norma provém do produto interno.

Por outro lado, a linearidade e a simetria nos dizem que

$$\begin{aligned}\langle \xi y - \vartheta x, \xi y - \vartheta x\rangle &= \langle \xi y, \xi y - \vartheta x\rangle - \langle \vartheta x, \xi y - \vartheta x\rangle \\ &= \langle \xi y, \xi y\rangle - \langle \xi y, \vartheta x\rangle - \langle \vartheta x, \xi y\rangle + \langle \vartheta x, \vartheta x\rangle \\ &= \xi^2\langle y, y\rangle - 2\xi\vartheta\langle x, y\rangle + \vartheta^2\langle x, y\rangle \\ &= \xi^2\|y\|^2 - 2\xi\vartheta\langle x, y\rangle + \vartheta^2\|x\|^2 \\ &= 2\xi^2\vartheta^2 - 2\xi\vartheta\langle x, y\rangle.\end{aligned}$$

Portanto, concluímos que

$$2\xi^2\vartheta^2 - 2\xi\vartheta\langle x, y\rangle \geq 0,$$

o que equivale a dizer que

$$2\xi\vartheta\langle x, y\rangle \leq \xi 2\vartheta^2 \quad \Longleftrightarrow \quad \langle x, y\rangle \leq \xi\vartheta.$$

Ou seja, $\langle x, y\rangle \leq \|x\|\|y\|$. Essa é uma desigualdade está, então, provada para quaisquer par de vetores x e y. Em particular, dado um par x, y, podemos aplicá-la ao par $-x, y$, e assim obtemos $-\langle x, y\rangle \leq \|x\|\|y\|$. Isso demonstra então que[4]

$$|\langle x, y\rangle \leq \|x\|\|y\|.$$

Examinando a demonstração, verificamos que a única maneira de termos uma igualdade é se tivermos $\langle \xi y - \vartheta x, \xi y - \vartheta x\rangle$, o que, pela positividade definida do produto interno nos diz que $\xi y - \vartheta x = 0$ e portanto x e y são colineares. Isso termina a prova. □

Agora, a desigualdade triangular para a função $\|x\| = \sqrt{\langle x, x\rangle}$ segue diretamente:

$$\begin{aligned}\|x + y\|^2 &= \langle x + y, x + y\rangle = \langle x, x + y\rangle + \langle y, x + y\rangle \\ &= \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle \\ &= \|x\|^2 + 2\langle x, y\rangle + \|y\|^2 \\ &\leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 \\ &= (\|x\| + \|y\|)^2.\end{aligned}$$

Tomando a raiz quadrada de ambos os lados obtemos finalmente que $\|x + y\| \leq \|x\| + \|y\|$. Isso demonstra que a função $\|x\| = \sqrt{\langle x, x\rangle}$ é de fato uma norma.

[4]Toda vez que a, b são números reais, com b positivo, e $a \leq b$ e $-a \leq b$ vale que $|a| \leq b$.

4.2.2 Ângulo entre vetores

Com a norma advinda do produto interno podemos definir o ângulo entre dois vetores.

Definição 4.2.2. *Seja E um espaço vetorial munido de um produto interno $\langle .,. \rangle$ e sejam $x, y \in E$. O ângulo entre x e y é o número real*

$$\arccos\left(\frac{\langle x, y\rangle}{\|x\|\|y\|}\right).$$

Observe que, pela desigualdade de Cauchy-Scwartz, temos que

$$\frac{\langle x, y\rangle}{\|x\|\|y\|} \in [-1, 1],$$

portanto esse quociente sempre é o cosseno de algum ângulo. A definição de ângulo entre vetores num espaço vetorial com produto interno pode ser reinterpretada como uma fórmula. De fato, se θ é o valor numérico do ângulo entre x e y então a definição diz que

$$\langle x, y\rangle = \|x\|\|y\| \cos\theta.$$

Ou seja, o cosseno do ângulo entre x e y é quanto o produto das normas difere do produto interno (em porcentagem, já que são números com valor absoluto menor do que 1).

Com a noção de ângulo bem definida, toda a linguagem da geometria passa a ter um significado preciso aqui. Dois vetores são *ortogonais* quando o ângulo entre eles é de 90º (note que isso acontece se, e somente se, o produto interno dá zero). Podemos dizer que os vetores formam um ângulo agudo, ou um ângulo obtuso entre si. Além disso, o ângulo dá zero ou 180º somente quando há igualdade na desigualdade de Cauchy-Scwartz, portanto quando os vetores são colineares.

4.3 Distância entre vetores

Seja E um espaço vetorial munido de uma norma $N : E \to \mathbb{R}$ qualquer (pode ser oriunda de um produto interno ou não). Podemos medir a distância entre vetores usando a norma. Sejam $x, y \in E$. A *distância* entre x e y (de acordo com a norma N) é o número

$$d_N = N(x - y).$$

A distância é usada como uma medida de quanto os pontos x e y diferem no espaço E. Na maioria das aplicações a distância é um elemento crucial para a construção de modelos.

Esse capítulo foi bastante abstrato, eu sei. Mas todo esforço sempre será recompensado. No próximo capítulo vamos estudar em detalhes um caso concreto: quando $E = \mathbb{R}^d$, e N é a norma que vem do produto interno euclideano. E vamos ver aplicações desse caso concreto. Mas antes vamos ver umas aplicações maneiras de se ter a habilidade de medir comprimento e ângulo de vetores.

Como calcular a distância do Pão de Açúcar ao Pelourinho

Vamos usar agora o material que aprendemos até aqui para responder a seguinte pergunta: como *calcular* a distância entre dois pontos na superfície da terra? Para resolver esse problema, vamos começar *modelando* a terra como uma esfera perfeita enfiada dentro do espaço $\mathbb{R}^3$ e vamos supor que o centro dessa esfera é a origem $O = (0, 0, 0)$. Cada ponto na superfície da terra é localizado por sua latitude θ e sua longitude λ. A latitude é o ângulo, em radianos, com relação ao equador. A longitude é o ângulo com relação ao meridiano central. Veja a Figura 4.2.

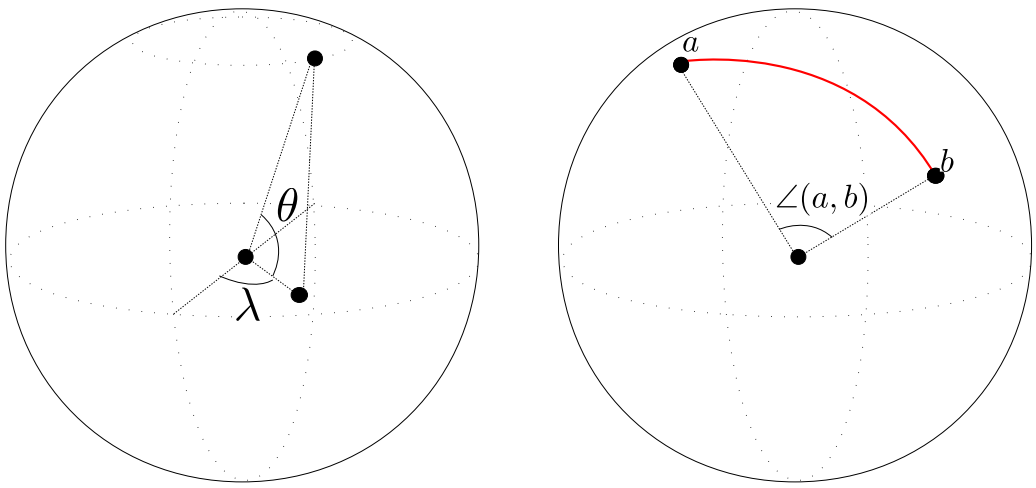

Figura 4.2: Medindo a distância entre dois pontos na superfície de uma esfera.

Sejam a, b dois pontos na superfície da terra (ou seja, dois pontos numa esfera de raio R). A distância entre eles na superfície da terra é o comprimento do menor arco de círculo

ligando a e b. Ora, o comprimento de um arco de círculo é $R\angle(a,b)$, lembrando que o raio da terra é $\approx 6367.5Km$. Então, o que resta saber para podermos determinar $R\angle(a,b)$ é como calcular o ângulo entre a e b, tendo apenas a informação da latitude e da longitude?

Bom, para calcular o ângulo entre a e b precisamos descobrir, a partir de suas respectivas latitude e longitude, suas coordenadas em $\mathbb{R}^3$. Tendo essa informação, podemos aplicar diretamente a definição de ângulo:

$$\angle(a,b) = \arccos\left(\frac{\langle a,b\rangle}{\|a\|\|b\|}\right).$$

E para obter, as coordenadas de um ponto qualquer $p=(x,y,z)$ em função de sua latitude e de sua longitude, basta olhar a Figura 4.3: a coordenada z é o cateto oposto ao ângulo θ num triângulo retângulo de hipotenusa R, e as coordenadas x e y são os catetos num triângulo retângulo de hipotenusa $R\cos\theta$. Logo, concluímos que se $p=(x,y,z)$ é um ponto

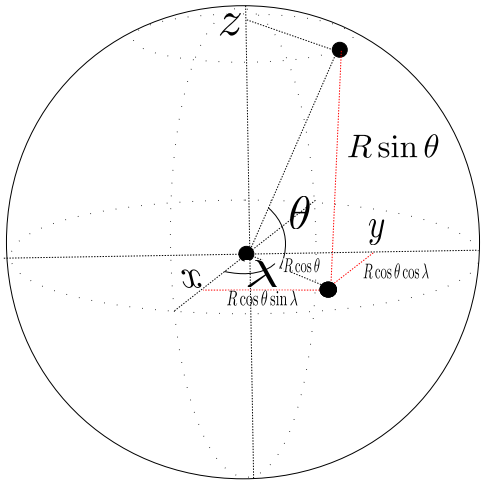

Figura 4.3: As coordenadas de $p=(x,y,z)$ em função de sua latitude e de sua longitude.

na superfície da terra com latitude θ e longitude λ então

$$x = R\cos\theta\cos\lambda, \quad y = R\cos\theta\sin\lambda \text{ e } z = R\sin\theta.$$

Vamos usar isso para calcular a distância entre o Pão de Açúcar e o Pelourinho. Usando o *Google maps*, você digita "Pão de Açúcar"clica em *O que há aqui* e ele te mostra a latitude e a longitude. Assim, encontramos $\theta = -22.9491^o$ e $\lambda = -43.154704^o$. Podemos escrever uma

função em **Julia** que recebe as coordenadas de latitude e longitude e retorna as coordenadas na superfície da terra.

function coord(θ,λ)
p=6367.5*[cos((θ*2*π)/360)*cos((λ*2*π)/360),cos((θ*2*π)/360)*sin((λ*2*π)/360),sin((θ*2*π)/360)]
return p
end

Observe que é preciso colocar os ângulos em radianos. Essa função diz que a posição do Pão de Açúcar é

$$a = (4277.495894296885, -4010.4770407316205, -2482.772448300052).$$

Para inserir uma letra grega em Julia basta digitar "barra theta"e dar tab. A latitude e a longitude do Pelourinho são $\theta = -12.971790$ e $\lambda = -38.508264$. Assim, as coordenadas do Pelourinho são

$$b = (4855.531027884479, -3863.407351310128, -1429.3209314545006).$$

Agora resta apenas escrever a função que dá ângulo (não esquece de dar using LinearAlgebra antes para ter acesso a função norma)

function ang(x,y)
α=acos((x'*y)/(norm(x)*norm(y)))
return α
end

A conta $R\angle(a, b)$ dá 1212.4149028241127Km.

Capítulo 5

Aplicações da norma euclideana

Este capítulo será inteiramente dedicado ao estudo da norma euclideana

$$\|x\| = \sqrt{\sum_{\ell=1}^{d} x_\ell^2}, \quad , x \in \mathbb{R}^d,$$

suas propriedades *finas* e aplicações. Em particular, vamos ver outros exemplos de normas advindas de outros produtos internos e como essas maneiras alternativas de medir a "magnitude"de um vetor se relacionam com a norma euclideana.

5.1 O valor rms

Uma das "desvantagens"da norma euclideana é que ela é ruim para uma análise estatística das entradas de um vetor. Por exemplo, a norma do vetor de uns $\mathbb{I} = (1, ..., 1) \in \mathbb{R}^d$ depende de d (de fato, $\|\mathbb{I}\| = \sqrt{d}$) enquanto que do ponto de vista estatístico esse vetor é o mesmo para qualquer valor de d.

Em outras palavras, imagine que busquemos uma noção estatística do que deve ser a magnitude de um vetor. Por exemplo, o valor "típico"das entradas de x. A norma euclideana não fornece isso. Precisamos de uma outra maneira de medir comprimentos. Uma forma de fazer isso é através do valor rms, que vamos definir a seguir.

5.1.1 O produto interno rms

O *produto interno* rms em $\mathbb{R}^d$ é a função

$$\begin{aligned} R : \mathbb{R}^d \times \mathbb{R}^d &\to \mathbb{R} \\ (x, y) \in \mathbb{R}^d \times \mathbb{R}^d &\mapsto R(x, y) \stackrel{\text{def}}{=} \frac{1}{\sqrt{d}} \sum_{\ell=1}^{d} x_\ell y_\ell, \end{aligned}$$

A função R é um tipo de produto euclideano com pesos[1] (no caso, os pesos são todos iguais). Em particular, R define um produto interno em $\mathbb{R}^d$.

Definição 5.1.1. *Seja $x \in \mathbb{R}^d$ o valor* rms *de x é a norma de x advinda do produto interno* rms*. Em outras palavras,*

$$\text{rms}(x) \stackrel{\text{def}}{=} R(x, x).$$

A sigla rms vem do inglês *root mean square value*, que pode ser traduzido livremente como "raiz do valor médio quadrático", ou ainda de modo mais criativo como "valor médio *raiz*"[2]. No entanto, vamos usar o termo em inglês mesmo. Como vimos no capítulo anterior, a função

$$\text{rms} : \mathbb{R}^d \to [0, +\infty)$$

é uma norma em $\mathbb{R}^d$. Da definição de R segue que

$$\text{rms}(x) = \frac{\|x\|}{\sqrt{d}} = \sqrt{\frac{\sum_{\ell=1}^{d} x_\ell^2}{d}}.$$

O valor rms de x tenta dizer qual o valor típico do módulo de uma entrada de x. Assim, por exemplo $\text{rms}(\mathbb{I}) = 1$, para qualquer $d \in \mathbb{N}$, o que é natural pois as entradas de $\mathbb{I}$ não apresentam nenhuma variação. Mais geralmente, qualquer vetor que tenha todas as entradas iguais vai ter o valor rms igual a entrada constate (que é o valor típico das entradas). Em outras palavras,

$$\text{rms}(c\mathbb{I}) = |c|,$$

o que é uma consequência imediata da homogeneidade da norma $\text{rms}(x)$. Observe que o nome "*root mean square value*"é bastante sugestivo: $\text{rms}(x)$ é a raiz quadrada do valor médio dos quadrados das entradas.

[1] Veja a lista de exercícios

[2] Ou seja: o valor médio que não é nutella...

5.2 A desigualdade de Chebyshev

A desigualdade de Chebyshev é um resultado fácil de demonstrar e cuja aplicabilidade supera, em muito, o esforço necessário para obtê-la.

Proposição 5.2.1 (Desigualdade de Chebyshev)**.** *Seja $x \in \mathbb{R}^d$ e seja $\varepsilon > 0$. Suponha que n das d entradas de x (logo, evidentemente estamos supondo $n < d$) tenham valor absoluto maior do que ε. Ou seja, existe um subconjunto $K \subset \{1, ..., d\}$, com n elementos, tal que que se $j \in K$ então $|x_j| \geq \varepsilon$. Então,*

$$\frac{n}{d} \leq \left(\frac{\text{rms}(x)}{\varepsilon}\right)^2.$$

Exemplo 5.2.2. *Seja $x \in \mathbb{R}^d$, e $\varepsilon = 3\,\text{rms}(x)$. Seja n tal que n dentre as entradas de x possuem valor absoluto maior do que ε. Então, a desigualdade de Chebyshev diz que*

$$\frac{n}{d} \leq \left(\frac{\text{rms}(x)}{\varepsilon}\right)^2 = \frac{1}{9}.$$

Como $1/9 \approx 0.1111$, concluímos que menos do que $11.2\,\%$ das entradas de x podem exceder 3 vezes o valor $\text{rms}(x)$. Esse exemplo ilustra bem como a desigualdade de Chebyshev corrobora a ideia de que $\text{rms}(x)$ é o valor absoluto típico de uma entrada de x, já que poucas entradas de x podem exceder por fator "grande"o valor $\text{rms}(x)$. Observe que se ao invés de 3, usarmos um número ainda maior, concluímos que a proporção de entradas que excedem $\text{rms}(x)$ por esse fator é ainda menor.

Demonstração da Desigualdade de Chebyshev. Seja $j \in K$. Então

$$x_j^2 \geq \varepsilon^2, \ \forall\, j \in K.$$

Se $j \notin K$ podemos dizer tão somente que $x_j^2 \geq 0$. Logo,

$$\|x\|^2 = x_1^2 + ... + x_d^2 \geq n\varepsilon^2.$$

Ou seja,

$$n \leq \frac{\|x\|^2}{\varepsilon^2}.$$

Como $\|x\|^2 = d\,\text{rms}(x)^2$, concluímos que

$$n \leq d\left(\frac{\text{rms}(x)}{\varepsilon}\right)^2,$$

de onde a desigualdade procurada segue diretamente. Isso completa a prova. □

5.3 Medindo distâncias em $\mathbb{R}^d$

O conceito geométrico de distância entre pontos encontra muitas aplicações imediatas quando olhamos os espaços euclideanos $\mathbb{R}^d$. Afinal, medir a distância entre dois pontos e medir o quanto esses pontos diferem um do outro, com relação a um certo aspecto. Isso leva a diversas maneiras deiferentes de se medir distâncias, através de normas diferentes.

Assim, por exemplo, podemos considerar a distância euclideana usual entre dois pontos:

$$d(x, y) \stackrel{\text{def}}{=} \|x - y\|, \ \forall \ x, y \in \mathbb{R}^d.$$

Ou podemos considerar a *raiz do desvio quadrático médio*, que é a distância induzida pela norma rms:

$$d_{\text{rms}}(x, y) \stackrel{\text{def}}{=} \text{rms}(x - y), \quad \forall \ x, y \in \mathbb{R}^d.$$

5.3.1 Aplicações

Distância de atributos

Suponha que $x, y \in \mathbb{R}^d$ representam atributos (medições médicas) de dois pacientes. Se $d(x, y)$ é pequena podemos dizer que os casos clínicos de x e y são parecidos. Essa observação pode ser usada, por exemplo, para criar algorítimos que auxiliem a solução de casos difíceis: o algorítimo pode fornecer casos semelhantes que possam revelar algum padrão, auxiliando na busca da solução. Uma pergunta para você, caro leitor: qual maneira de medir distâncias você usaria?

Erro de previsão de temperatura

Suponha que $x \in \mathbb{R}^d$ represente uma série temporal que forneça a temperatura medida hora-a-hora em alguma localização. Suponha que $\hat{x} \in \mathbb{R}^d$ seja um outro vetor que forneça predições da temperatura na mesma localização. A distância entre x e $\hat{x}$ é o *erro de previsão*. Observe que nesse caso é muito mais interessante usar a distância d_{rms}, pois o valor rms é uma medida estatística que não depende tanto de d: ou seja, o que importa é que a maioria

das previsões seja acertada.

Dissimilaridade de documentos

Suponha que $x, y \in \mathbb{R}^d$ representem os histogramas de ocorrência de palavras em dois documentos. Nesse caso, a distância entre x e y representa uma medida da dissimilaridade entre os dois documentos. É de se esperar que a dissimilaridade seja pequena quando os dois documentos tem o mesmo assunto, mesmo tema, mesmo autor, e que seja maior se não.

Vizinho mais próximo

Suponha que $y_1, ..., y_n \in \mathbb{R}^d$ seja um conjunto finito de pontos, e seja $x \in \mathbb{R}^d$ um outro ponto qualquer. Dizemos que y_ℓ é o ponto mais próximo de x (dentre os pontos $\{y_1, ..., y_n\}$) se

$$d(x, y_\ell) \leq d(x, y_j), \ \ \forall \ j \in \{1, ..., n\}.$$

5.4 Desvio Padrão

Nesta seção vamos ver mais uma medida estatística que podemos associar a um vetor $x \in \mathbb{R}^d$. A ideia básica é olhar a médias das entradas e depois olhar como cada entrada se desvia da média, e tomar o valor rms desses desvios.

5.4.1 A média e o vetor centrado

A medida estatística mais simples associada a um vetor de $\mathbb{R}^d$ é o valor médio das suas entradas.

Definição 5.4.1 (Média)**.** *Dado $x \in \mathbb{R}^d$ a* média *de x é o número $\mu(x) \stackrel{\text{def}}{=} \frac{1}{d} \sum_{\ell=1}^{d} x_\ell$.*

Dado um vetor $x \in \mathbb{R}^d$, podemos construir a partir dele um novo vetor com cada nova entrada igual à discrepância da respectiva entrada com relação à média.

Definição 5.4.2 (Versão centrada)**.** *Seja $x \in \mathbb{R}^d$. A versão centrada de x é o vetor $x^c \stackrel{\text{def}}{=} x - \mu(x)\mathbb{I}$.*

Observe que a j-ésima entrada da versão centrada de x é $x_j - \mu(x)$. Ou seja, a j-ésima entrada de x^c é o quanto x_j difere da média $\mu(x)$.

Do ponto de vista da teoria geral dos espaços vetoriais, ao considerar a versão centrada de x estamos definindo um *operador linear* em $\mathbb{R}^d$, ou ainda um *campo de vetores linear*, conceitos esses que vamos ver com carinho mais à frente. No entanto, a ideia básica é de uma transformação. Estamos transformando o vetor x num novo vetor, criando assim a sua versão centrada. Para entender melhor um novo conceito é sempre bom estudar suas propriedades. A seguir vamos ver uma propriedade geométrica da versão centrada de x.

Lema 5.4.3. *Sejam $x, y \in \mathbb{R}^d$. Então, x^c e y^c estão alinhados se e somente se x e y estão alinhados.*

Demonstração. Observe que dois vetores estão alinhados se, e somente se, eles múltiplos um do outro. Formalmente, isso quer dizer que existe um número real $\lambda > 0$ tal que $x = \lambda y$.

Suponha que x e y estejam alinhados. Então, $x = \lambda y$ para algum número real $\lambda > 0$. Deixo como exercício para o leitor verificar que isso implica que $\mu(x) = \lambda\mu(y)$[3]. Então

$$x^c = x - \mu(x)\mathbb{I} = \lambda y - \lambda\mu(y)\mathbb{I} = \lambda(y - \mu(y)\mathbb{I}) = \lambda y^c.$$

Logo x^c e y^c estão alinhados (pelo mesmo fator de alinhamento de x e y). □

5.4.2 O desvio padrão

Agora se a gente toma o valor rms da versão centrada de x o que vamos obter vai ser o valor "esperado"de quanto uma entrada de x se desvia da média.

Definição 5.4.4 (Desvio padrão)**.** *Seja $x \in \mathbb{R}^d$. O* desvio padrão *de x é $\sigma(x) \stackrel{\text{def}}{=} \text{rms}(x^c)$.*

Lembrando da definição do valor rms, podemos escrever algumas fórmulas clássicas para o desvio padrão.

[3]Veja o exercício *Linearidade da média*

- $\sigma(x) = \frac{\|x-\mu(x)\mathbb{I}\|}{\sqrt{d}}$.
- $\sigma(x) = \frac{\sqrt{\sum_{\ell=1}^{d}(x_\ell-\mu(x))^2}}{\sqrt{d}}$.

Vamos ver agora algumas propriedades do desvio padrão.

Lema 5.4.5. *Para todo $x \in \mathbb{R}^d$ e todo número real λ valem as seguintes identidades*

1. $\sigma(x + \lambda\mathbb{I}) = \sigma(x)$;
2. $\sigma(\lambda x) = |\lambda|\sigma(x)$.

Esse lema pode ser provado de duas formas diferentes: uma bruta e outra elegante. A forma bruta consiste em usar a expressão explícita de $\sigma(x)$ em termos das coordenadas. A forma elegante consiste em usar as *propriedade estruturais* da função $x \mapsto x^c$. Essa função é um exemplo do que vamos ver mais tarde nesse livro: uma transformação linear! Com eu sou aficionado por elegância, vamos postergar a prova do Lema 5.4.5. No entanto, considero um importante exercício para você, caro leitor, parar e tentar fazer a demonstração na força bruta. Isso vai te permitir, no futuro, apreciar mais a elegância.

Falando em elegância, o desvio padrão se relaciona com a média e com o valor rms através dessa fórmula linda

$$\mathrm{rms}(x)^2 = \mu(x)^2 + \sigma(x)^2, \quad \forall\, x \in \mathbb{R}^d. \tag{5.1}$$

Prova da Fórmula 5.1. Pela definição do desvio padrão (veja a primeira fórmula clássica) temos que

$$d\sigma(x)^2 = \|x - \mu(x)\|^2. \tag{5.2}$$

Usando o Teorema de Pitágoras (veja o exercício *O teorema de Pitágoras em espaços de Hilbert*) temos

$$\|x - \mu(x)\mathbb{I}\|^2 = \|x\|^2 - 2\langle x, \mu(x)\mathbb{I}\rangle + d\mu(x)^2,$$

uma vez que $\|\mu(x)\mathbb{I}\|^2 = d\mu(x)^2$.

Agora lembre que tomar o produto interno pelo vetor de uns dá a soma das entradas: $\langle x, \mathbb{I}\rangle = \sum_{\ell=1}^{d} x_\ell$. Portanto,

$$\mu(x) = \frac{1}{d}\langle x, \mathbb{I}\rangle \quad \Longrightarrow \quad d\mu(x) = \langle x, \mathbb{I}\rangle.$$

Combinando isso com a homogeneidade do produto interno, $\langle x, \mu(x)\mathbb{I}\rangle = \mu(x)\langle x, \mathbb{I}\rangle = d\mu(x)^2$. Tudo isso nos leva a concluir que

$$\|x - \mu(x)\mathbb{I}\|^2 = \|x\|^2 - 2d\mu(x)^2 + d\mu(x)^2 = \|x\|^2 - d\mu(x)^2.$$

Usando isso na igualdade (5.2) concluímos que

$$\|x\|^2 = d\mu(x)^2 + d\sigma(x)^2 \quad \Longrightarrow \quad \frac{\|x\|^2}{d} = \mu(x)^2 + \sigma(x)^2.$$

Como $\mathrm{rms}(x)^2 = \|x\|^2/d$, isso nos dá a fórmula procurada. □

Retorno médio e risco

Suponha que $x \in \mathbb{R}^d$ represente a série temporal dos retornos de um investimento, em porcentagem, em d intervalos de tempo (por exemplo d meses, ou d dias). Então $\mu(x)$ fornece o retorno médio do investimento. Em finanças, $\mu(x)$ é carinhosamente chamado apenas de *retorno*. O desvio padrão é uma medida de quão variável o retorno é de tempos em tempos, após algum intervalo de tempo. Em outras palavras (supondo medições de retorno diárias), $\sigma(x)$ fornece o quanto o retorno tipicamente se desloca da média. Em finanças ele é conhecido como *risco* do investimento. Por exemplo, na figura a seguir plotamos 4 investimentos, olhando seus retornos mensais em 10 meses.

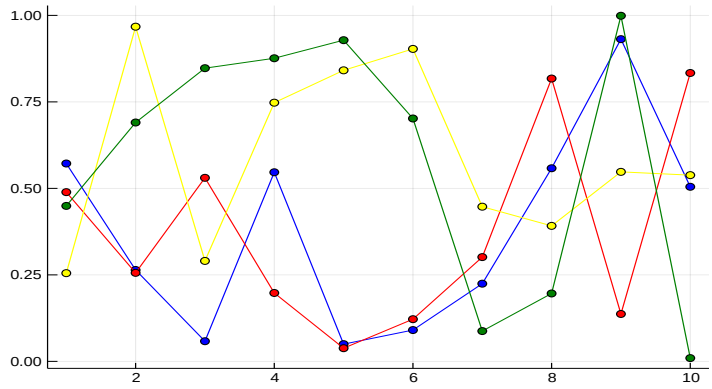

Figura 5.1: Cada gráfico representa um vetor em $\mathbb{R}^d$ que fornece o retorno de um dado investimento num período de 10 meses.

Um investimento desejável deve ter retorno alto e risco baixo. Múltiplos investimentos podem ser comparados num gráfico *retorno-risco*, como na figura 5.2 abaixo, onde são com-

parados os 4 investimentos acima. No eixo horizontal marcamos o retorno do investimento e no eixo vertical o risco.

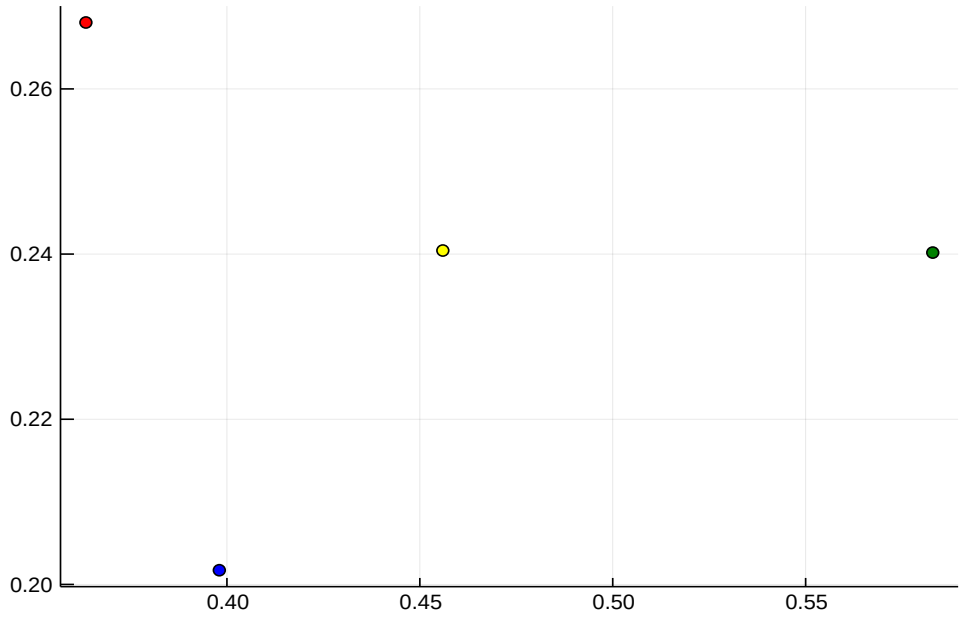

Figura 5.2: Olhando esse gráfico, qual o melhor investimento?

Abaixo os comandos em **Julia** que usei para fazer essas figuras.

Figura 5.1

a=rand(10)

b=rand(10)

c=rand(10)

d=rand(10)

plot(a, marker=:circle, color=:blue, legend=false)

plot!(b, marker=:circle, color=:red, legend=false)

plot!(c, marker=:circle, color=:yellow, legend=false)

plot!(d, marker=:circle, color=:green, legend=false)

savefig("abcd.pdf")

Figura 5.2

plot((avg(a),σ(a)), marker=:circle, color=:blue, legend=false)

plot!((avg(b),σ(b)), marker=:circle, color=:red, legend=false)

plot!((avg(c),σ(c)), marker=:circle, color=:yellow, legend=false)
plot!((avg(d),σ(d)), marker=:circle, color=:green, legend=false)
savefig("returnrisk.pdf")

Série temporal de temperatura

Suponha que $x \in \mathbb{R}^d$ seja a série temporal das temperaturas médias diárias ao longo de d dias (por exemplo, um ano para fixar ideias) de uma localização particular. Então $\mu(x)$ fornece a temperatura média dessa localização e o desvio padrão é uma medida do quanto a temperatura diária se desviou da média. Se a gente lembrar das aulas de geografia do ensino básico, é de se esperar que a temperatura média seja alta e o desvio padrão baixo em uma região tropical, e o oposto em uma região de alta latitude.

Uma coisa divertida é calcular essas medidas (norma, média, desvio padrão, etc...) com Julia e plotar o resultado. Você pode ensinar Julia a calcular a média de um vetor, definindo uma função:

```
function avg(x)
a=sum(x)/length(x)
return a
end
```

A norma de um vetor, se você usa o pacote LinearAlgebra (não esquece de botar using LinearAlgebra). Usando using Plots a gente pode plotar num gráfico as entradas do vetor e ver como elas oscilam. Além disso, dá para plotar no mesmo gráfico a média e a oscilação da média (o desvio padrão), como na figura abaixo.

A figura 5.3 foi obtida com o comando

```
plot(x, color=:blue, marker= :circle)
plot!(avg(x)*ones(length(x)), color= :red)
plot!((avg(x)+σ(x))*ones(length(x)), color= :yellow)
plot!((avg(x)-σ(x))*ones(length(x)), color= :yellow)
```

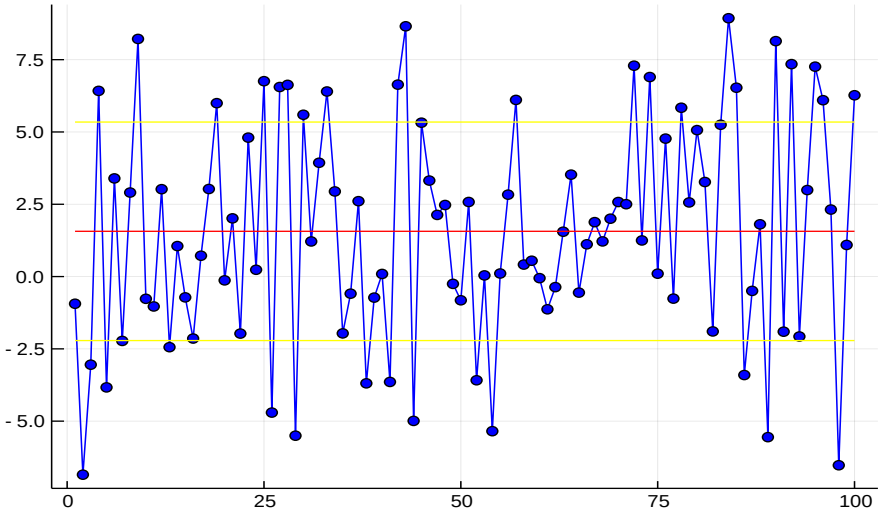

Figura 5.3: O gráfico azul representa a medição diária de temperatura média em 100 dias numa cidade. O gráfico vermelho é a média. Os gráficos amarelos são os vetores $x \pm \sigma(x)\mathbb{I}$. Se fosse um gráfico de temperaturas, qual cidade você acha que seria?

```
plot!(legend=false)
savefig("avgplot.pdf")
```

Desigualdade de Chebyshev com desvio padrão

Dado $x \in \mathbb{R}^d$, aplicando a desigualdade de Chebyshev ao vetor x^c concluímos que se n das d entradas de x satisfazem $|x_i - \mu(x)| \geq \varepsilon$ então

$$\frac{n}{d} \leq \left(\frac{\sigma(x)}{\varepsilon}\right)^2.$$

Como exemplo de aplicação suponha que $x \in \mathbb{R}^d$ represente a série temporal de um investimento com retorno 8.6% e risco 6%. Na figura a seguir está representado um exemplo específico com $x \in \mathbb{R}^{11}$.

Observe que quando uma coordenada de x é negativa, necessariamente o descolamento absoluto em relação à média, $|x_i - \mu(x)|$, é maior do que $|\mu(x)|$. Como no nosso exemplo $\mu(x) = 8.6\%$, a desigualdade de Chebyshev nos diz que o número máximo possível n de coordenadas negativas de x, portanto o número máximo de *perdas* no investimento, satisfaz

$$\frac{n}{d} \leq \left(\frac{6}{8.6}\right)^2 = 48.6\%.$$

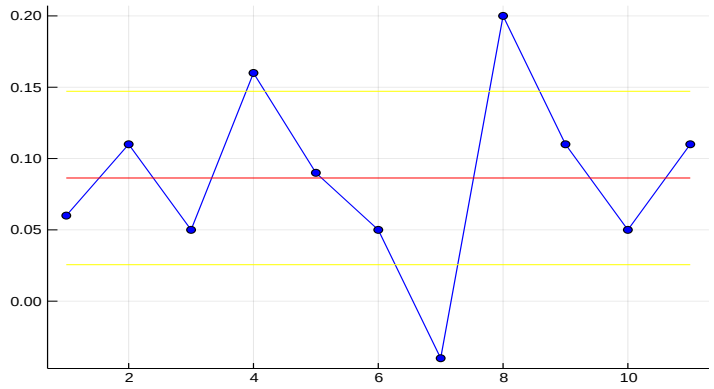

Figura 5.4: Retornos mensais de um investimento ao longo de 8 meses. O retorno médio é de 8.6% e o risco é de 6%.

Ou seja, nesse caso o investimento tem um máximo de 48.6% de perdas. Fica claro também que quanto menor o *risco* $\sigma(x)$ menor o número de perdas.

5.4.3 Coeficiente de correlação

Sejam $x, y \in \mathbb{R}^d$. O *coeficiente de correlação* entre x e y é o número

$$\rho(x,y) \stackrel{\text{def}}{=} \frac{\langle x^c, y^c \rangle}{\|x^c\|\|y^c\|}.$$

A desigualdade de Cauchy-Schwartz (veja a Proposição 4.2.1) implica que $\rho(x,y) \in [-1,1]$, e que $\rho(x,y) \in \{1,-1\}$ se, e somente se, x^c e y^c estão alinhados. O coeficiente de correlação diz como as entradas nos dois vetores variam conjuntamente. Um coeficiente de correlação alto, perto de 1 significa que tipicamente quando x está acima da média y também está. Ou seja, as entradas de x e y estão conjuntamente acima da média muitas vezes. Isso leva a esperar, por exemplo, que dois vetores de medições de temperatura em localidades muito próximas sejam muito correlacionados.

O desvio padrão da soma

Vamos usar o coeficiente de correlação para deduzir uma fórmula para desvio padrão da soma de dois vetores.

Lema 5.4.6 (Desvio padrão da soma)**.** *Sejam $x, y \in \mathbb{R}^d$. Então*

$$\sigma(x+y) = \sqrt{\sigma(x)^2 + 2\rho(x,y)\sigma(x)\sigma(y) + \sigma(y)^2}.$$

Demonstração. Nessa demonstração vamos usar a *lei dos cossenos em espaços de Hilbert*[4]: seja E um espaço vetorial com produto interno e seja $\|.\|$ o norma induzida pelo produto interno. Então

$$\|x+y\|^2 = \|x\|^2 + 2\langle x, y\rangle + \|y\|^2.$$

Aqui, o nosso espaço vetorial é $\mathbb{R}^d$ e o produto interno é o produto interno euclideano. Sejam $x, y \in \mathbb{R}^d$. Pela definição do desvio padrão,

$$d\sigma(x+y)^2 = \|x^c + y^c\|^2 = \|x^c\|^2 + 2\langle x^c, y^c\rangle + \|y^c\|^2.$$

Pela definição do coeficiente de correlação, $\langle x^c, y^c\rangle = \rho(x,y)\|x^c\|\|y^c\|$. Como a definição do desvio padrão nos dá $\|x^c\| = \sqrt{d}\sigma(x)$ e, similarmente, $\|y^c\| = \sqrt{d}\sigma(y)$, concluímos que

$$\begin{aligned} d\sigma(x+y)^2 &= d\sigma(x)^2 + 2d\rho(x,y)\sigma(x)\sigma(y) + d\sigma(y)^2 \\ \therefore \sigma(x+y)^2 &= \sigma(x)^2 + 2\rho(x,y)\sigma(x)\sigma(y) + \sigma(y)^2, \end{aligned}$$

o que evidentemente dá a fórmula anunciada. □

Hedging investimentos

Com a teoria descrita nesse capítulo podemos *explicar* uma estratégia de "pulverização"de investimentos muito utilizada no mercado financeiro e que visa reduzir o risco. Vamos ver como combinando aportes em dois portfólios com mesmo retorno e mesmo risco produz um novo portfólio com retorno idêntico, mas com risco menor.

Para ver isso, Suponhamos que $x, y \in \mathbb{R}^d$ representem as séries temporais de retornos de dois portfólios de investimento de tal forma que $\mu(x) = \mu(y)$ (os investimentos têm o mesmo retorno) e $\sigma(x) = \sigma(y)$ (os investimentos têm o mesmo risco). Para fazer contas e simplificar a notação vamos dar um nome mais fácil ao retorno comum $m = \mu(x) = \mu(y)$ (m de média),

[4]Veja o exercício *o teorema de Pitágoras em espaços de Hilbert*

e $r = \sigma(x) = \sigma(y)$ (e r de risco). Considere o vetor obtido pela combinação convexa de x e y com coeficientes $1/2, 1/2$; ou seja

$$z = \frac{x}{2} + \frac{y}{2}.$$

No contexto de finanças o vetor z corresponde a um *blending* dos investimentos x e y. Em bom português é um investimento misturado entre partes de x e partes de y. Como as entradas de z são $z_i = x_i/2 + y_i/2$ retorno do investimento z é[5]

$$\mu(z) = \frac{1}{d}\sum_{i=1}^{d} \frac{x_i}{2} + \frac{y_i}{2} = \frac{1}{2}\left(\frac{1}{d}\sum_{i=1}^{d} x_i + \frac{1}{d}\sum_{i=1}^{d} y_i\right) = \frac{1}{2}\left(\mu(x) + \mu(y)\right) = m.$$

Isso prova que o investimento combinado de metade de x e metade de y dá o mesmo retorno que davam x e y. E quanto ao risco? Ora, o risco do investimento z pode ser determinando usando a fórmula para o desvio padrão da soma que acabamos de ver. Seja $c = \rho(x, y)$ o coeficiente de correlação entre x e y. Usando o Lema 5.4.5 vemos que $\sigma(\frac{x}{2}) = \frac{\sigma(x)}{2}$ e $\sigma(\frac{y}{2}) = \frac{\sigma(y)}{2}$. Logo,

$$\sigma(z) = \sqrt{\frac{r^2}{4} + \frac{2cr^2}{4} + \frac{r^2}{4}} = \frac{\sqrt{2r^2 + 2cr^2}}{2} = r\frac{\sqrt{2}}{2}\sqrt{1+c}. \tag{5.3}$$

O leitor pode ter se perguntado: se x e y têm os mesmos retornos e risco então são o mesmo investimento? Bom, a resposta é não. Uma maneira eficaz de diferenciar os dois investimentos é olhar o coeficiente de correlação. Eles são iguais apenas se o coeficiente de correlação for 1. No entanto, o mais importante é que a expressão $r\frac{\sqrt{2}}{2}\sqrt{1+c}$ é sempre menor do que r. De fato, seu maior valor é atingindo quando o coeficiente de correlação é igual a 1, e nesse caso

$$\sigma(z) = r\frac{1}{2}\sqrt{2}\sqrt{2} = r.$$

Ou seja, os investimentos tem o mesmo risco. Agora, se os investimentos são não correlacionados, então

$$\sigma(z) = r\frac{\sqrt{2}}{2} \approx r \times 0.707.$$

Ou seja, o risco do investimento *blended* é quase 30% menor. No caso ideal em que os investimentos são negativamente correlacionados ($c = -1$), o risco seria zero. No entanto, uma análise mais conceitual da expressão (5.3) explica completamente porque fazer *blending*

[5]Mais a frente vamos ver que μ é um *funcional linear* e com esse conceito o cálculo do retorno é imediato: $\mu(\frac{x+y}{2}) = \frac{\mu(x)+\mu(y)}{2} = m$.

sempre reduz o risco. A função $c \mapsto \frac{\sqrt{2}}{2}\sqrt{1+c}$ é crescente.[6] Como o coeficiente de correlação só pode variar de -1 até 1, seu mínimo é zero (quando $c = -1$) e seu maximo é 1 (quando $c = 1$). Qualquer que seja o coeficiente de correlação, a expressão (5.3) nos diz que o risco

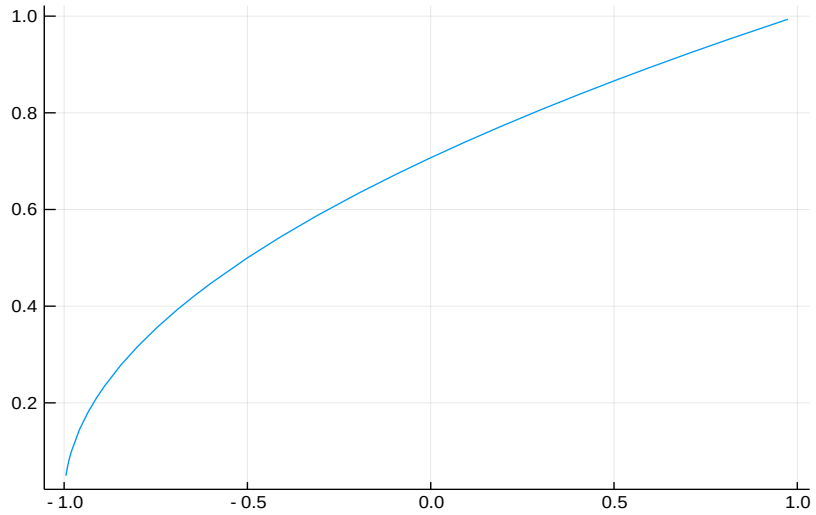

Figura 5.5: Gráfico da função $c \mapsto \frac{\sqrt{2}}{2}\sqrt{1+c}$.

do investimento z é sempre o risco comum r de x e y multiplicado por um número entre 0 e 1.

Misturar investimentos pouco, ou negativamente correlacionados é que se chama o mercado financeiro de *hedging*. Como vimos, quanto mais negativamente correlacionados forem os investimentos, menor é o risco do investimento combinado.

5.5 Julia

5.5.1 A super calculadora Julia

Você pode calcular a norma de um vetor em Julia. Aqui um programinha que faz isso:

```
function normae(x,d)
n=sqrt(sum(x.^2)) println(n) end
```

[6] Se você não gosta muito de estimativas abstratas, pode ir no **Julia** e plotar o gráfico, e ver o resultado na figura 5.5

E, por exemplo, normae([1,1,1,1],4)=2. Mas você não precisa escrever um programa, pois em Julia a função norma euclideana já vem pronta no pacote LinearAlgebra. Você tem que dizer ao Julia que quer o usar o pacote

```
using LinearAlgebra
```

e aí é só sair calculando normas a torto e a direito com norm(x). Por exemplo, se se $x \in \mathbb{R}^5 0$ é um vetor aleatório cujas entradas representam porcentagem, você pode visualisar x e calcular sua norma.

```
x=rand(50)
plot(x,marker = :circle, legend = false, grid = false)
norm(x)
```

Capítulo 6

Aglutinação de dados (*Clustering data*)

Este capítulo é inteiramente dedicado a uma aplicação. Ele é especialmente simpático pois mostra como o material que vimos até aqui (o qual apenas arranhou de leve a superfície do imenso *iceberg* da álgebra linear) serve para para formalizar e explicar uma técnica extremamente útil. Vamos aprender como e **porque funciona** o algoritmo $k-means$, muito utilizado em aplicações que envolvem Ciência de Dados. Trata-se de uma ferramente heurística desenhada para resolver (de forma aproximada, porém satisfatória) o seguinte problema: *dado um conjunto enorme de dados, é possível particioná-lo em poucas aglutinações nas quais todos os dados são (de alguma forma) "parecidos"?*

6.1 Formalização do problema

Seja E um espaço vetorial munido de um produto interno $\langle , \rangle$ e seja $\|.\|$ a norma oriunda desse produto interno. Para as aplicações em Ciência de Dados, vamos usar sempre $E = \mathbb{R}^d$ munido do produto interno euclideano.

Considere um conjunto de n vetores $X = \{x_1, ..., x_N\} \subset E$. Queremos encontrar uma partição de X em k subconjuntos, dois a dois disjuntos, e cuja união seja todo o X.

Definição 6.1.1 (Agrupamento)**.** *Um* agrupamento *(ou* partição*) de X em k pedaços é uma família de subconjuntos $G_1 \subset X$,...,$G_k \subset X$, satisfazendo*

1. $i \neq j \implies G_i \cap G_j = \emptyset$
2. $\cup_{\ell=1}^{k} G_\ell = X$

Uma outra maneira de interpretar a noção de agrupamento é imaginar os pontos de X dispostos no espaço e pensar que o agrupamento é uma maneira de colorir os elementos de X com k cores diferentes. Assim, podemos condensar toda a informação de um agrupamento na *função de coloração* $c : [1, N] \to [1, k]$ definida por

$$c(\ell) = j \iff x_\ell \in G_j.$$

Ou seja, se você enumera as cores de 1 até k então a função de coloração aplicada ao número 5 por exemplo te diz o número da cor que agrupamento atribui ao 5º vetor de X. Pensar no agrupamento por meio da função de coloração é um pouco mais rebuscado tecnicamente porém extremamente útil conceitualmente. Observe que uma função qualquer $c : \{1, ..., N\} \to \{1, ..., k\}$ é uma função de coloração de algum agrupamento se, e somente se, c é sobrejetiva.

6.1.1 Representantes do grupo

Como a ideia que temos à primeira vista do que deve ser um agrupamento significativo é que os vetores dentro de um mesmo grupo devem ser todos "parecidos"entre si, é natural pensar que eles devem ser todos parecidos a um mesmo "representante"do grupo.

Definição 6.1.2 (Representantes)**.** *Seja $G_1, ..., G_k$ um agrupamento do conjunto X. Dada um conjunto qualquer $Z = \{z_1, ..., z_k\} \subset E$ com k elementos dizemos que o vetor z_ℓ é o representante do grupo G_ℓ.*

O leitor pode achar estranho a flexibilidade na definição. Basicamente estamos dizendo que qualquer conjunto com k elementos é uma escolha de representantes! Mas, você deve ter em mente que o objetivo do algoritmo $k-means$ é buscar a melhor escolha de representantes,

aquela que faz com que todos os vetores no grupo G_ℓ sejam parecidos com o representante z_ℓ. Mais precisamente, dentre as infinitas escolhas possíveis de representantes vamos buscar aquela para a qual

$$\|x_\ell - z_{c(\ell)}\|$$

seja o menor possível.[1] Como a gente quer que todos esses números sejam pequenos, vamos atribuir um número a escolha dos grupos (ou a função de coloração $c : \{1, ..., N\} \to \{1, ..., k\}$) e a escolha dos representantes.

Definição 6.1.3 (Qualidade do agrupamento)**.** *Seja $G_1, ..., G_k$ um agrupamento de X, traduzido pela função de coloração $c : \{1, ..., N\} \to \{1, ..., k\}$ com represantes $Z = \{z_1, ..., z_k\}$. A* qualidade do agrupamento *é o número*

$$J(c, Z) \stackrel{\text{def}}{=} \frac{1}{N} \sum_{\ell=1}^{N} \|x_\ell - z_{c(\ell)}\|^2.$$

6.1.2 Enunciado do problema

Formalmente, podemos agora traduzir matematicamente o problema do agrupamento de dados:

Problema 1 (Agrupamento de dados)**.** *Dado um subconjunto finito $X \subset E$ de N elementos, encontrar um agrupamento de X com função de coloração $c : \{1, ..., N\} \to \{1, ..., k\}$ e representantes $Z = \{z_1, ..., z_k\}$ para o qual a qualidade $J(c, Z)$ seja a menor possível.*

Observe que no enunciado a quantidade k de grupos é um dado, não algo a ser determinado. Iremos discutir mais sobre essa sutileza no fim do capítulo.

Um outro detalhe importante sobre o enunciado é que o objetivo é encontrar o agrupamento **a partir** dos dados. Ou seja, a ideia é que o agrupamento está escondido pela estrutura do conjunto de dados, e queremos usar a matemática para desmembrar essa estrutura interna e extrair dela o agrupamento.

[1] Lembre que $c(\ell) \in \{1, ..., k\}$ é o número do grupo ao qual x_ℓ pertence.

6.2 Motivação

Antes de prosseguir sobre como abordar o problema e como tentar resolvê-lo, vamos discutir ilustrar algumas situações nas quais o problema do agrupamento de dados aparece.

Exemplo 6.2.1 (Estudos sobre o Câncer)**.** *Suponha que um grupo de pesquisadores da Universidade Federal Fluminense estuda o comportamento de um tipo de câncer ao longo de sua evolução por 4 estágios. Essa é uma situação hipotética, porém com forte conexão com a realidade. Suponha que os cientistas tenham a sua disposição um conjunto enorme de dados médicos de pacientes de todo o Brasil. Como se trata de uma base de dados enorme, a informação de qual estágio de cada paciente não está disponível. Espera-se que pacientes em um mesmo estágio do câncer devem ter características médicas parecidas. Nessa situação, uma solução para o problema do agrupamento (como por exemplo o algoritmo $k-means$) de dados pode ser usada "adivinhar"o estágio no qual cada paciente se encontra. Mais do que isso, ainda que a informação esteja disponível comparar a resposta do algorítimo $k-means$ com a realidade também pode ser uma maneira de obter* insights *sobre o câncer!*

O leitor interessado em uma aplicação com dados do mundo real pode consultar `https://azul.netlify.com/2018/08/11/prouni-clustering/` no qual o estudante de economia da UFF, Pedro Cavalcante, aplica o $k-means$ para identificar cursos do Prouni e, em especial, para jogar luz nas anomalias associadas aos cursos de medicina. Nesse outro texto `https://azul.netlify.com/2018/08/19/problemas-clustering-k-means/`, o Pedro apresenta uma bela discussão sobre a sutileza da escolha do número k de grupos no agrupamento.

6.3 O algorítimo $k-means$

Descrito em breves palavras o $k-means$ funciona assim: dado o conjunto X, você dá um chute inicial Z_1 para o conjunto de representantes.[2] Com ele você define o primeiro agrupamento (ou a coloração c_1); ele é escolhido de forma a minimizar a distância aos representantes (essa é a parte fácil do algorítimo). Em seguida, você atualiza a definição

[2]Observe que, como a nossa definição de representantes é absurdamente flexível, você nem precisa conhecer o agrupamento para dar um chute dos representantes.

dos representantes: o novo conjunto de representantes Z_2 vai ser obtido tomando-se os centroides dos grupos definidos pela coloração c_1. Vamos demonstrar que fazendo-se isso, sempre teremos $J(c_1, Z_2) < J(c_1, Z_1)$: ou seja a nova escolha de representantes melhora a qualidade do agrupamento! Assim, com os novos representantes, redefinimos a coloração da mesma forma e continuamos. O algorítimo "converge"porque a cada etapa reduzimos a qualidade. Então podemos impor que ele pare no momento em que a qualidade for pequena o bastante (o quanto é esse pequeno o bastante depende, claro, da aplicação em cena).

6.3.1 O passo 1

Vamos nos colocar nas hipóteses do Problema 1. Então, seja E um espaço de Hilbert (ou seja, um espaço vetorial munido de um produto interno $\langle x, y\rangle$ e $\|x\| = \sqrt{\langle x, x\rangle}$ a norma advinda do produto interno). Seja $X = \{x_1, ..., x_N\} \subset E$ um conjunto finito de vetores (o nosso conjunto de dados).

Vamos descrever um procedimento iterativo de atribuição de coloração e representantes que, a cada etapa, reduz a qualidade do agrupamento. Na etapa 0 escolhemos aleatoriamente um conjunto $Z_1 = \{z_1^1, ..., z_k^1\}$ de representantes. Esse é o nosso chute inicial.

O passo inicial do algorítimo $k-means$ consiste em escolher uma coloração que minimize a qualidade, dentre todas as escolhas possíveis de coloração.

Lema 6.3.1 (Unicidade da coloração ótima)**.** *Existe uma única coloração* $c_1 : \{1, ..., N\} \to \{1, ..., k\}$ *tal que*

$$J(c_1, Z_1) \leq J(c, Z_1)$$

para toda função sobrejetiva $c : \{1, ..., N\} \to \{1, ..., k\}$.

Demonstração. Esse é um problema de minimização finito, por isso é fácil de resolver. De fato, observe que existe apenas um número finito de funções sobrejetivas de $\{1, ..., N\}$ sobre $\{1, ..., k\}$. Isso pode ser visto com um argumento elegante de combinatória do ensino médio: pelo princípio multiplicativo existem apenas k^N funções de $\{1, ..., N\}$ sobre $\{1, ..., k\}$, e portanto dentre essas aquelas que são sobrejetivas certamente estão em menor número. Havendo

apenas um número finito de colorações a escolher, podemos (pelo menos conceitualmente) testá-las uma a uma até achar aquela que minimiza a qualidade. □

Após o passo 1 temos uma partição de X, representada pela coloração $c_1 : \{1, ..., N\} \to \{1, ..., k\}$ e um conjunto de representantes $J(c_1, Z_1)$. No próximo passa, vamos redefinir a conjunto de representantes e com isso reduzir *efetivamente* a qualidade do agrupamento.

6.3.2 Passo 2

E escolha dos novos representantes é fornecida pelo teorema a seguir.

Teorema 6.3.2. *Sejam $X = \{x_1, ..., x_N\} \subset E$ um conjunto finito e seja $G_1, ..., G_k$ um agrupamento de X com função de coloração $c : \{1, ..., N\} \to \{1, ..., k\}$. Para cada grupo G_ℓ seja u_ℓ o seu centroide, ou seja*[3][4]

$$u_\ell = \frac{1}{|G_\ell|} \sum_{j \in c(\ell)} x_j.$$

Considere $Z = \{z_1, ..., z_k\} \subset E$ um conjunto qualquer com k elementos. Então, a qualidade do conjunto de representantes $\mathcal{U} \stackrel{\text{def}}{=} \{u_1, ..., u_k\}$ é estritamente menor do que a qualidade de Z. Em outras palavras,

$$J(c, \mathcal{U}) < J(c, Z).$$

De acordo com esse teorema, se a gente troca o chute inicial Z de representantes pelos centroides u_ℓ dos grupos (que definimos usando o chute inicial de representantes) conseguimos diminuir a qualidade do agrupamento. O algorítimo $k-means$ consiste em repetir os passos 1 e 2, em cada etapa reduzindo um pouco mais a qualidade. A aplicação particular em tela vai dizer até quando vamos querer reduzir a qualidade.

No restante dessa seção vamos nos dedicar a demonstração do Teorema 6.3.2, que explica porque o algorítimo funciona.

[3]Lembre dada uma função qualquer $f : A \to B$ ente dois conjuntos A e B, dado $b \in B$, a pré-imagem de b por f é o conjunto $f^{-1}(b) \stackrel{\text{def}}{=} \{a \in A; f(a) = b\}$. Em particular, f é sobrejetiva se toda pré-imagem é não vazia e é injetiva se toda pré-imagem não-vazia for um conjunto unitário.

[4]Para A um conjunto finito $|A|$ denota a quantidade de elementos de A.

O centroide é o ponto mais próximo

Como acontece muitas vezes em matemática, o passo fundamental na solução de um problema consiste em resolver um caso muito particular. Aqui, vamos resolver o caso extremo e aparentemente inútil em que temos apenas um grupo.

Proposição 6.3.3 (Lema do Centroide)**.** *Sejam $Y = \{y_1, ..., y_N\} \subset E$ e $u = \frac{1}{N}\sum_{j=1}^{N} y_j$ o centroide (ou a media) do conjunto Y. Seja $z \in E$ qualquer. Considere as quantidades*

$$J(u) = \sum_{j=1}^{N} \|y_j - u\|^2 \quad e \quad J(z) = \sum_{j=1}^{N} \|y_j - z\|^2.$$

Se $z \neq u$ então $J(u) < J(z)$.

Ou seja, estamos dizendo que dentre todos os pontos do espaço E, aquele que realiza o menor valor da soma dos quadrados das distâncias aos pontos do conjunto finito Y é o centroide de Y. Além disso, essa minimização é única: qualquer outro ponto diferente do centroide fornece um valor maior para a soma dos quadrados das distâncias.

Demonstração. Vamos usar a *lei dos cossenos em espaços de Hilbert:*[5]

$$\|a + b\|^2 = \|a\|^2 + 2\langle a, b\rangle + \|b\|^2 \tag{6.1}$$

Seja $j \in \{1, ..., N\}$ qualquer. A equação (6.1) com $a = y_j - u$ e $b = -(z - u)$ nos leva a

$$\|y_j - z\|^2 = \|y_j - u - (z - u)\|^2 = \|y_j - u\|^2 - 2\langle y_j - u, z - u\rangle + \|z - u\|^2.$$

Logo,

$$\begin{aligned} J(z) = \sum_{j=1}^{N} \|y_j - z\|^2 &= \sum_{j=1}^{N} \big(\|y_j - u\|^2 - 2\langle y_j - u, z - u\rangle\big) + N\|z - u\| \\ &= \sum_{j=1}^{N} \|y_j - u\| - 2\sum_{j=1}^{N}\langle y_j - u, z - u\rangle + N\|z - u\|. \end{aligned}$$

Pela linearidade do produto interno,

$$\sum_{j=1}^{N}\langle y_j - u, z - u\rangle = \langle\sum_{j=1}^{N}(y_j - u), z - u\rangle.$$

[5]Veja o exercício "O teorema de Pitágoras em espaços de Hilbert"da lista de exercícios.

O leitor questionador deve ter se perguntado: de onde raios saiu essa escolha marota de $a = y_j - u$ e $b = -(z - u)$. Pois aqui vem a resposta: a definição de u implica que

$$\sum_{j=1}^{N}(y_j - u) = \sum_{j=1}^{N} y_j - Nu = 0.$$

Concluímos assim que

$$J(z) = \sum_{j=1}^{N} \|y_j - u\| + N\|z - u\| = J(u) + N\|z - u\|.$$

Portanto, se $z \neq u$ então $\|z - u\| > 0$ e segue que $J(z) > J(u)$. □

Prova do Teorema 6.3.2

Seja $J(z_\ell) = \sum_{j \in c^{-1}(\ell)} \|x_j - z_\ell\|^2$. Então, $J(c, Z) = \frac{1}{N}\sum_{\ell=1}^{k} J(z_\ell)$. Similarmente, definimos $J(u_\ell) = \sum_{j \in c^{-1}(\ell)} \|x_j - u_\ell\|^2$ e com isso temos $J(c,\mathcal{U}) = \frac{1}{N}\sum_{\ell=1}^{k} J(u_\ell)$. A Proposição 6.3.3 aplicada ao conjunto G_ℓ diz que $J(u_\ell) < J(z_\ell)$, de onde deduzimos que $J(c,\mathcal{U}) < J(c, Z)$. □

Algorítimo $k - means$

Dada uma lista de vetores $\{x_1, ..., x_n\} \subset \mathbb{R}^d$ e uma lista inicial $\{z_1, ..., z_k\} \subset \mathbb{R}^d$ de k representantes, repita os passos a seguir até a convergência
Passo 1 Escolha uma função de coloração com a melhor qualidade possível.
Passo 2 Redefina os representantes como centroides dos grupos.

6.3.3 Aplicações

Algorítimo de recomendação

O $k-means$ pode ser usada para o desenho de um mecanismo de recomendação de músicas, ou filmes, por usuários de uma plataforma de *streaming*, como Spotify ou Netflix. Suponha que para cada usuário da plataforma criemos um vetor com tantas entradas quantas forem as músicas na plataforma (vamos ficar com o Spotify para fixar as ideias) e tal que a entrada j fornece o número de vezes (dentro um período de tempo, como alguns meses por exemplo)

que um usuário escutou a música j. Tipicamente, esse vai ser um vetor esparso (com muitas entradas nulas), o que é ótimo computacionalmente. Ao rodar o algorítimo $k - means$ podemos agrupar os usuários da plataforma por "gostos musicais"semelhantes. Assim dado um usuário qualquer, a podemos sugerir para ele as músicas tipicamente escutadas pelas pessoas com gosto musical semelhante a ele (os membros do mesmo grupo), mas que ele nunca escutou.

Adivinhar as entradas, ou como roubar no Sudoku

Uma aplicação do $k - means$, que tem um sabor semelhante a anterior, porém com um espectro mais amplos de aplicações é usá-lo para completar entradas não conhecidas de alguns elementos de um conjunto grande de dados. Por exemplo, suponha que um conjunto de dados $X = \{x^1, ..., x^k\} \subset \mathbb{R}^d$ de tamanho d represente atributos dos alunos que prestaram o ENEM no último ano (idade, cor, raça, religião, renda *per capita* da família, etc...), obtidas através do questionário socioeconômico. Suponha ainda que uma certa proporção desse conjunto de dados, digamos 6%, corresponde a estudantes que não responderam completamente o questionário. É possível adivinhar o que eles responderiam?

Podemos propor uma solução para esse problema usando o algorítimo $k - means$. Funciona assim: você considera o subconjunto de X formado apenas pelos vetores "completos". Você "roda"o $k-means$ nesse subconjunto, obtendo assim um conjunto $\{z^1, ..., z^k\}$ de representantes. Suponha, por exemplo, que $x^j \in X$ tenha algumas de suas entradas desconhecidas (correspondendo, portanto, à questões não-respondidas do questionário). Usando as entradas conhecidas de x^j determinamos o represante mais próximo. Basta considerar o subconjunto $\mathcal{K} \subset \{1, ..., d\}$ correspondente às entradas conhecidas e tomar $\ell \in \{1, ..., k\}$ que minimize

$$\sum_{i \in \mathcal{K}} (x_i - z_i^j)^2.$$

Então, você "chuta"a entrada faltante de x como sendo igual a correspondente do seu representante mais próximo.

Implementação

Vamos agora discutir com um pouco implementar em **Julia** o algorítimo $k - menas$.

Vamos começar analisando o problema da aglutinação de dados com um conjunto pequeno, "listado à mão", de dados planares (ou seja, vetores de $\mathbb{R}^2$).

```
x = [ [0,1], [1,0], [-1,1],[0.9,-0.9],[1,-1],[1.1,1.07],[0.51,-0.23],[0.9,-0.8]]
```

Visualizamos esse conjunto de dados com o código

```
scatter([c[1] for c in x], [c[2] for c in x])
plot!(lims = (-2, 2.1), size = (500,500), legend = false)
```

que produz como resultado

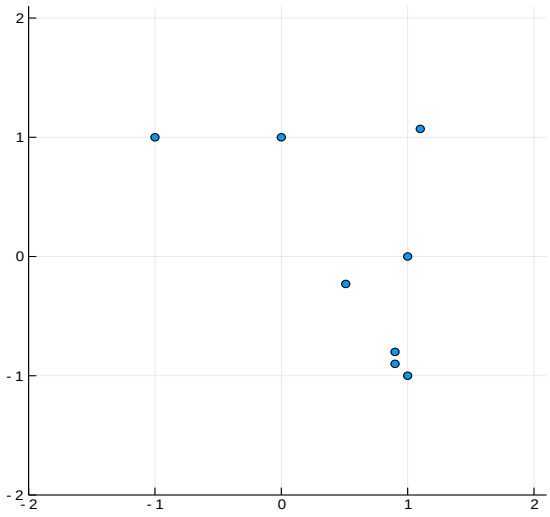

Figura 6.1: Dataset

Vamos definir aleatoriamente uma função de coloração, chutando que o nosso conjunto de dados vai se dividir bem em 3 grupos. Em seguida, plotamos dividido em três cores.

```
colorac=[rand(1:3) for i in 1:8]
k=3
N=length(x)
grps = [[x[i] for i=1:N if colorac[i] == j] for j=1:k]
scatter([c[1] for c in grps[1]], [c[2] for c in grps[1]])
scatter!([c[1] for c in grps[2]], [c[2] for c in grps[2]])
scatter!([c[1] for c in grps[3]], [c[2] for c in grps[3]])
plot!(legend = false, grid = false, size = (500,500),
xlims = (-1.5,2.5), ylims = (-2,2))
```

A qualidade desse agrupamento se calcula pela fórmula

```
Jclust(x,reps,colorac) = avg( [norm(x[i]-reps[colorac[i]])^2 for i=1:length(x)])
```

Para o agrupamento da figura 6.2 gerado de forma aleatória a qualidade dá 1.5247375. Agora vamos rodar o algorítimo $k-means$ no nosso conjunto de dados e ver qual coloração ele retorna. Essa implementação eu retirei do *Julia companion* do Stephen Boyd e do Lieven Vandenberghe, deixando inclusive os comentários explicativos de cada etapa do algorítimo.

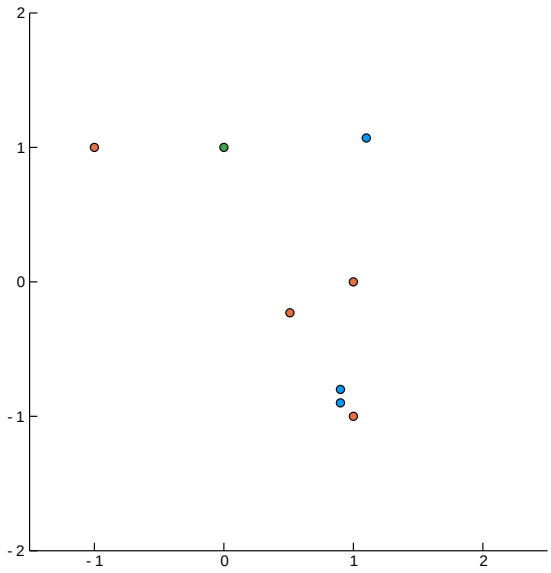

Figura 6.2: Data set agrupado aleatoriamente.

```
function kmeans(x, k; maxiters = 100, tol = 1e-5)


N = length(x)
n = length(x[1])
distances = zeros(N) # usado para armazenar a distância de cada
#ponto ao seu representante mais próximo.
reps = [zeros(n) for j=1:k] # usado para armazenar os representantes.
# 'assignment'(=coloração) é um vetor de ℝ^N com entradas inteiras variando de 1 a k.
# A coloração inicial é escolhida aleatoriamente.
assignment = [ rand(1:k) for i in 1:N ]
Jprevious = Inf # used in stopping condition
for iter = 1:maxiters
# O representante do agrupamento j é o centroide da galera no agrupamento j.
for j = 1:k
group = [i for i=1:N if assignment[i] == j]
reps[j] = sum(x[group]) / length(group);
end;
# Para cada x[i], encontra a distância ao representante mais próximo
# e a 'cor' associada a ele.
for i = 1:N
(distances[i], assignment[i]) =
findmin([norm(x[i] - reps[j]) for j = 1:k])
end;
# Calcula a qualidade do agrupamento.
J = norm(distances)^2 / N
# Show progress and terminate if J stopped decreasing.
println("Iteration ", iter, ": Jclust = ", J, ".")
if iter > 1 && abs(J - Jprevious) < tol * J
return assignment, reps
end
Jprevious = J
end
end
```

Agora que criamos uma função que roda $k-means$ num conjunto de dados vamos aplicar ela ao nosso conjunto gerado a mão.

```
assignment, reps = kmeans(x, 3)
Iteration 1: Jclust = 0.879982638888889.
Iteration 2: Jclust = 0.19690000000000002.
Iteration 3: Jclust = 0.19690000000000002.
([3, 1, 3, 2, 2, 1, 2, 2], ArrayFloat64,1[[1.05, 0.535], [0.8275, -0.7325], [-0.5, 1.0]])
```

Repare que a qualidade final do agrupamento é 0.1969, muito melhor do que a qualidade do agrupamento aleatório. O resultado é

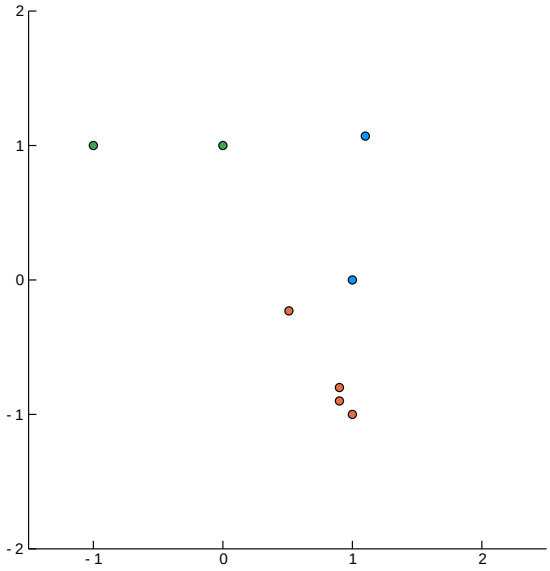

Figura 6.3: Data set agrupado com $k - menas$.

Para terminar, vamos gerar aleatoriamente um conjunto planar de 100 vetores.

O $k - means$ retorna

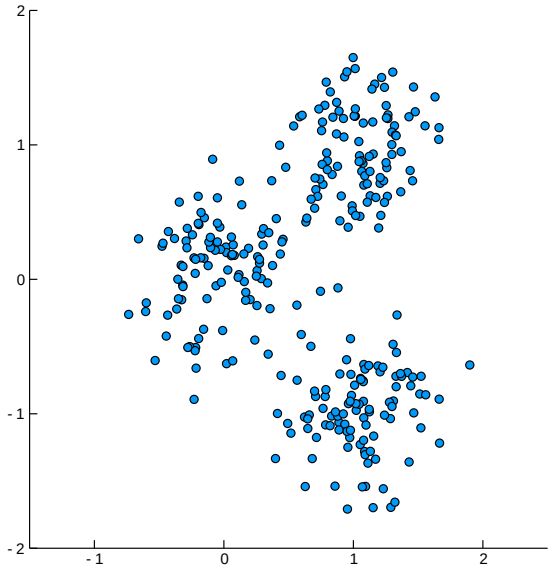

Figura 6.4: Data set aleatório

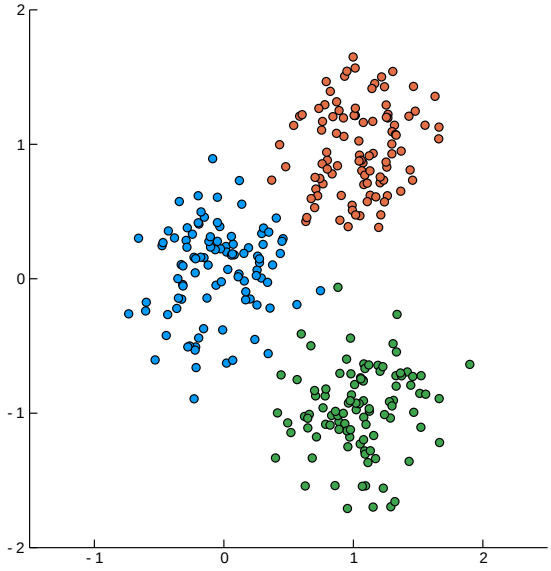

Figura 6.5: Data set aleatório aglutinado

Capítulo 7

Base e dimensão

Até aqui vimos como olhar vetores, medir comprimento e ângulo tem inúmeras aplicações fantásticas. No entanto, o material que cobrimos até aqui é apenas o térreo do imenso e imponente edifício da álgebra linear e de suas fantásticas aplicações. Por isso, nesse capítulo vamos aprofundar um pouco o estudo da estrutura algébrica dos espaços vetoriais. Esse estudo nos levará a um conceito fundamental: a dimensão de um espaço vetorial.

7.1 Subespaços

Definição 7.1.1. *Seja E um espaço vetorial. Seja $F \subset E$ um subconjunto de E. Dizemos que F é um* subespaço *de E se satisfaz as seguintes propriedades*

1. $0 \in F$;
2. *se* $u, v \in F$ *então* $u + v \in F$;
3. *se* $u \in F$ *e* $\lambda \in \mathbb{R}$ *então* $\lambda u \in F$.

Em palavras: um subespaço é um subconjunto que contém o vetor nulo e que é estável pelas operações de soma de vetores e multiplicação por número real. Mais geralmente, se

$v_1, ..., v_n \in F$ e $\lambda_1, ..., \lambda_n \in \mathbb{R}$ então

$$\sum_{j=1}^{n} \lambda_j v_j \in F,$$

ou seja: subespaços são estáveis por combinações lineares.

Exemplo 7.1.2. *Em $\mathbb{R}^3$, o conjunto P dos vetores da forma $v = (0, x, y)$ é um subespaço vetorial. Com efeito, evidentemente $0 \in P$. Vejamos que P é estável com respeito a soma de vetores. Se $a = (0, x_1, y_1), b = (0, x_2, y_2) \in P$ então*

$$a + b = (0, x_1 + x_2, y_1 + y_2) \in P.$$

Analogamente, se $a = (0, x, y) \in P$ e $\lambda \in \mathbb{R}$ então $\lambda a = (0, \lambda x, \lambda y) \in P$.

Exemplo 7.1.3. *Em $C^0(\mathbb{R}, \mathbb{R})$ considere $F \subset C^0(\mathbb{R}, \mathbb{R})$ o subconjunto formado pelas funções diferenciáveis. Como a soma e o produto de funções diferenciáveis é diferenciável, e como a função constante nula segue que F é um subespaço de $C^0(\mathbb{R}, \mathbb{R})$.*

Observação 7.1.4. *Existem funções contínuas (bizarras) que não possuem derivada em ponto nenhum! O gráfico delas é um fractal com inclinação vertical em todos os pontos. Num certo sentido é possível dizer que a maioria das funções contínuas não possui derivada em ponto nenhum.*

Exemplo 7.1.5. *Seja E um espaço vetorial. Seja $v \in E$ um vetor. O conjunto $F = \{\lambda v; \lambda \in \mathbb{R}\}$ é um espaço subespaço vetorial. Com efeito, fazendo $\lambda = 0$ vemos que $0 \in F$. Se $\alpha v, \beta v \in F$ então, pela propriedade distributiva, $\alpha v + \beta v = (\alpha + \beta)v \in F$. Analogamente, se $\alpha v \in F$ e $\beta \in \mathbb{R}$ então $\beta(\alpha v) = (\beta\alpha)v \in F$.*

Definição 7.1.6. *Sejam X e Y subespaços vetoriais de um espaço vetorial E. A* soma *de X e Y é o conjunto de todos os vetores $x + y$, tais que $x \in X$ e $y \in Y$. Em símbolos:*

$$X + Y := \{v \in E; v = x + y,\ x \in X, y \in Y\}.$$

Equações diferenciais

O exemplo a seguir ilustra como a subida em abstração pode ser vantajosa uma vez que abre uma janela para geometria entrar no estudo de equações diferenciais, objetos fundamentais nas aplicações da matemática.

Exemplo 7.1.7 (Soluções de uma equação diferencial linear)**.** *Considere a equação diferencial*

$$x''(t) + x(t) = 0.$$

O conjunto E das funções que satisfazem essa equação é um espaço vetorial. Com efeito, seus elementos são as funções $x : \mathbb{R} \to \mathbb{R}$ que somadas a sua derivada segunda dão sempre zero. Já vimos como somar funções e multiplicá-las por números reais, e já vimos que tais operações sempre satisfazem os axiomas de espaço vetorial. Portanto, para termos certeza de que E é de fato um espaço vetorial basta checar que E é estável pelas operações de soma e multiplicação por um número real. Com efeito, se $x, y \in E$ temos que

$$\frac{d^2}{dt^2}(x(t) + y(t)) + (x(t) + y(t)) = \frac{d^2}{dt^2}x(t) + x(t) + \frac{d^2}{dt^2}y(t) + y(t) = 0,$$

pois a derivada da soma é a soma das derivadas. Além disso,

$$\frac{d^2}{dt^2}(\lambda x(t)) + \lambda x(t) = \lambda(\frac{d^2}{dt^2}x(t) + x(t)) = 0.$$

Nas subseções a seguir iremos dissecar as estruturas intrínsecas de um espaço vetorial. Esta estrutura é dada por propriedades que são consequências lógicas dos axiomas de espaço de vetorial.

7.2 Revisando combinações lineares

Num espaço vetorial podemos somar vetores e "esticá-los", multiplicando-os por números reais. Além disso, o axioma de associatividade diz que podemos somar três vetores, quatro, etc... e mais geralmente qualquer número n de vetores. Assim, dados n vetores $v_1, ..., v_n \in E$, podemos esticar cada um deles e depois somar tudo.

Definição 7.2.1. *Dados n vetores $v_1, ..., v_n$ e n números reais $\lambda_1, ..., \lambda_n$, o vetor*

$$v = \lambda_1 v_1 + ... + \lambda_n v_n$$

é chamado uma combinação linear *dos vetores $v_1, ..., v_n$.*

Observação 7.2.2. *Por questões de estética vamos utilizar neste livro a notação de somatório*

$$\sum_{i=1}^{n} \lambda_i = \lambda_1 + \lambda_2 + ... + \lambda_n,$$

que torna muito concisa e elegante expressões como combinações lineares, que serão recorrentes ao longo do livro.

Exemplo 7.2.3. *1. Sejam $e_1 = (1, 0)$ e $e_2 = (0, 1)$ vetores em $\mathbb{R}^2$. Então $v = (3, 7)$ é combinação linear de e_1 e e_2. Com efeito,*

$$3e_1 + 7e_2 = (3, 0) + (0, 7) = (3, 7).$$

2. Mais geralmente, todo vetor $v = (x_1, ..., x_d) \in \mathbb{R}^d$ é combinação linear dos vetores $e_1 = (1, 0, 0, ..., 0), e_2 = (0, 1, 0, 0, ..., 0), ..., e_d = (0, 0, ..., 1)$, pois

$$\sum_{j=1}^{n} x_j e_j = (x_1, 0, ...0) + ... + (0, 0, ..., x_d) = (x_1, ..., x_d).$$

3. Em $\mathbb{R}^3$, $v = (3 + 5\pi, 6, 31)$ é combinação linear dos vetores $v_1 = (1, 2, -4)$ e $v_2 = (\pi, 0, 7)$ pois

$$3v_1 + 5v_2 = (3, 6, -4) + (5\pi, 0, 35) = (3 + 5\pi, 6, 31).$$

4. No espaço $C^0(\mathbb{R}, \mathbb{R})$ das funções contínuas, sejam $f(t) = t^2$ e $g(t) = \sin(t)$. Então

$$h(t) = 3t^2 + 9\sin(t)$$

é combinação linear de f e g.

No lema abaixo, estendemos a propriedade de distributividade para um número qualquer de vetores.

Lema 7.2.4. *Seja E um espaço vetorial e sejam $\lambda \in \mathbb{R}$ e $v_1, ..., v_m \in E$. Então*

$$\lambda \sum_{j=1}^{m} v_j = \sum_{j=1}^{m} \lambda v_j.$$

Demonstração. Fazemos a prova por indução em m. O caso $m = 2$ é o axioma de distributividade. Suponhamos verdade para $m = k$ e vamos provar que é verdade para $m = k + 1$.

Então, queremos verificar que

$$\lambda \sum_{j=1}^{k+1} v_j = \sum_{j=1}^{k+1} \lambda v_j.$$

Vejamos,

$$\lambda \sum_{j=1}^{k+1} v_j = \lambda \left(\sum_{j=1}^{k} v_j + v_{k+1} \right).$$

Pela distributividade, temos que

$$\lambda \left(\sum_{j=1}^{k} v_j + v_{k+1} \right) = \lambda \sum_{j=1}^{k} v_j + \lambda v_{k+1}.$$

Pela hipótese de indução, podemos reescrever a primeira parcela

$$\lambda \sum_{j=1}^{k} v_j = \sum_{j=1}^{k} \lambda v_j,$$

e portanto

$$\lambda \sum_{j=1}^{k+1} v_j = \sum_{j=1}^{k+1} \lambda v_j,$$

como queria demonstrar. □

7.2.1 Subespaço gerado

Sejam E um espaço vetorial e $X = \{v_1, ..., v_m\}$. O *subespaço gerado* por X é o conjunto de todas as combinações lineares possíveis entre os elementos de X. Em símbolos,

$$\mathcal{S}(X) = \{\sum_{j=1}^{m} \alpha_j v_j; \alpha_1, ..., \alpha_m \in \mathbb{R}\}$$

Lema 7.2.5. *$\mathcal{S}(X)$ é um subespaço*

Demonstração. Para ver que $0 \in \mathcal{S}(X)$ basta tomar $\alpha_1 = ...\alpha_m = 0$. Se $u = \sum_{j=1}^{m} \alpha_j v_j, v = \sum_{j=1}^{m} \beta_j v_j \in \mathcal{S}(X)$, então

$$u + v = \sum_{j=1}^{m} \alpha_j v_j + \beta_j v_j = \sum_{j=1}^{m} (\alpha_j + \beta_j) v_j \in \mathcal{S}(X).$$

Analogamente, se $\lambda \in \mathbb{R}$ e $v = \sum_{j=1}^{m} \alpha_j v_j$ então

$$\lambda v = \lambda \sum_{j=1}^{m} \alpha_j v_j = \sum_{j=1}^{m} (\lambda \alpha_j) v_j \in \mathcal{S}(X).$$

Portanto, $\mathcal{S}(X)$ é estável por soma de vetores e multiplicação por escalar, logo é um subespaço, como queríamos demonstrar. □

Observe que se $F \subset E$ é um subespaço, e $v \in F$ então a reta gerada por v está contida em F, pois esta reta é o conjunto de todos os vetores múltiplos de v. Em outras palavras, como a reta gerada por v consiste nos vetores da forma λv, com $\lambda \in \mathbb{R}$, e como F é um subespaço vetorial, segue que esta reta está inteiramente contida em F.

Além disso, se v_1 e v_2 pertencem a F, e a_1 é um elemento da reta gerada por v_1 e b é um elemento da reta gerada por v_2 então $a + b \in F$, pois F é um subespaço vetorial.

Mais geralmente, se $F \subset E$ é um subespaço e $X = \{v_1, ..., v_m\} \subset F$ então, como cada $x_i \in F$, temos que dado qualquer $\lambda_i \in \mathbb{R}$ vale que $\lambda_i v_i \in F$, para todo $i = 1, ..., m$. Portanto, como F é estável por soma de vetores, segue que

$$\sum_{i=1}^{m} \lambda_i v_i \in F.$$

Acabamos de demonstrar o seguinte

Lema 7.2.6. *Seja E um espaço vetorial. Se $F \subset E$ é um subespaço, e $X = \{v_1, ..., v_m\} \subset F$ então $\mathcal{S}(X) \subset F$.*

Definição 7.2.7. *Se $\mathcal{S}(X) = E$ dizemos que X é um conjunto de* geradores *do espaço vetorial E.*

Exemplo 7.2.8. *No exemplo 7.2.3 vimos que $e_1 = (1, ...0), ..., e_d = (0, ..., 1)$ é um conjunto de geradores do espaço $\mathbb{R}^d$.*

7.3 Dependência Linear; Base e Dimensão

Considere os vetores $u = (1, 2)$ e $v = (3, 6)$ em $\mathbb{R}^2$. Note que $v = 3u$, portanto v é combinação linear de u. Assim, podemos dizer que v pertence ao subespaço gerado por u. Além disso, o subespaço gerado por u e v coincide com o subespaço gerado apenas por u. Com efeito, se $a = \alpha u + \beta v \in \mathcal{S}(u, v)$ é um elemento do subespaço gerado por u e v então podemos escrever

$$a = \alpha u + 3\beta u = (\alpha + 3\beta)u \in \mathcal{S}(u).$$

Ou seja mostramos com isso que $\mathcal{S}(u,v) \subset \mathcal{S}(u)$. Como evidentemente, temos que $\mathcal{S}(u) \subset \mathcal{S}(u,v)$, já que toda combinação linear só com u é uma combinação linear com u e v, concluímos que

$$\mathcal{S}(u) = \mathcal{S}(u,v).$$

Em palavras, mostramos que o subespaço gerado por u e por v coincide com o subespaço gerado apenas por u como afirmado.

Geometricamente, podemos enxergar esse fato da seguinte maneira: v já pertence à reta gerada por u, logo v é "irrelevante"na construção do subespaço gerado por u e v.

Vejamos mais um exemplo. Considere os vetores $u = (1,0,0)$, $v = (0,3,0)$, $w = (2,0,0)$ e $z = (\pi, -5, 0)$ em $\mathbb{R}^3$. Podemos ver, como antes, que $w = 2u$ e que $z = \pi u - \frac{5}{3}v$. Em palavras, w e z são combinações lineares de u e v. Usando isso, vamos mostrar que

$$\mathcal{S}(u,v,w,z) = \mathcal{S}(u,v).$$

Com efeito, sempre temos que $\mathcal{S}(u,v) \subset \mathcal{S}(u,v,w,z)$. Para mostrar que estes conjuntos são iguais, precisamos mostrar que a inclusão contrária $\mathcal{S}(u,v,w,z) \subset \mathcal{S}(u,v)$ é verdadeira. No entanto, isto equivale a demonstrar que toda combinação linear entre os vetores u, v, w e z pode ser reescrita como uma combinação linear apenas entre os vetores u e v.

E exatamente isto que passaremos a demonstrar agora. Então, considere $x = \alpha u + \beta v + \gamma w + \lambda z \in \mathcal{S}(u,v,w,z)$ uma combinação linear entre u, v, w e z. Como $w = 2u$ e $z = \pi u - \frac{5}{3}v$, podemos escrever

$$x = \alpha u + \beta v + 2\gamma u + \lambda(\pi u - \frac{5}{3}v) = (\alpha + 2\gamma + \lambda\pi)u + (\beta - \frac{5\lambda}{3})v \in \mathcal{S}(u,v).$$

O mesmo fenômeno se repetiu aqui: w estava já contido na reta gerada por u enquanto z estava já contido no plano gerado por u e v. Assim, w e z são "irrelevantes"na construção do subespaço gerado por pelos 4 vetores u, v, w e z, e este subespaço é apenas o plano gerado por u e v (veja Figura 7.1). Quando isso acontece, dizemos que o conjunto $\{u,v,w,z\} \in \mathbb{R}^3$ é *linearmente dependente*. Mais geralmente:

Definição 7.3.1. *Seja E um espaço vetorial. Seja $X = \{v_1, ..., v_m\} \subset E$. Dizemos que X é* linearmente dependente *(LD) se pelo menos um de seus elementos pode ser escrito como*

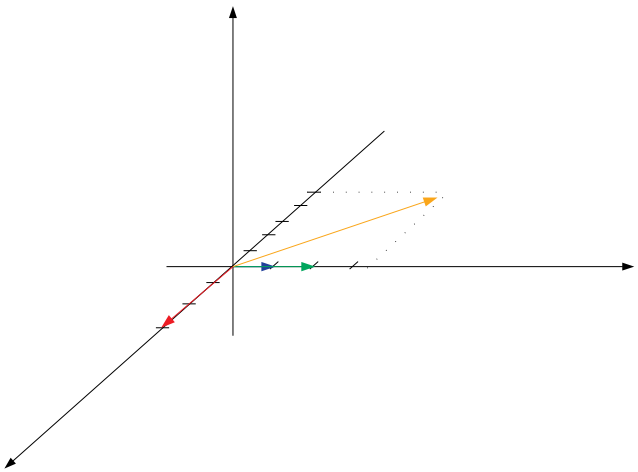

Figura 7.1: Os vetores amarelo e verde são redundantes na construção do subespaço gerado.

combinação linear dos demais. Caso X não seja linearmente dependente, dizemos que X é linearmente independente *(LI).*

Ou seja, um subconjunto de um espaço vetorial é linearmente independente se nenhum de seus elementos pode ser escrito como combinação linear dos demais.

Exemplo 7.3.2. *1. Em $\mathbb{R}^2$, o conjunto $X = \{(1,2),(3,6)\}$ é LD.*

2. Em $\mathbb{R}^4$ o conjunto $X = \{(1,0,0,0),(2,0,0,0),(0,3,0,0),(\pi,4,0,0)\}$ é LD.

3. Em $\mathbb{R}^2$, um conjunto $X = \{u,v\}$ é LD se, e somente se, v é um múltiplo de u.

Exemplo 7.3.3. *Em $\mathbb{R}^3$ os vetores $e_1 = (1,0,0)$, $e_2 = (0,1,0)$ e $e_3 = (0,0,1)$ formam um subconjunto LI. De fato, qualquer combinação entre e_2 e e_3 por exemplo terá a primeira coordenada nula, logo nunca poderá ser igual a e_1. Analogamente, vemos que qualquer combinação entre dois deles nunca será igual ao terceiro.*

Observe que para decidir se um conjunto de vetores é LD ou LI precisamos testar todas as possibilidades de combinação linear entre alguns dos vetores. Caso alguma delas seja igual a um vetor do conjunto, teremos um conjunto LD. Caso contrário, teremos um conjunto LI. Computacionalmente, parece muito complicado decidir se um conjunto de vetores num espaço vetorial é LD ou LI. O teorema a seguir soluciona essa dificuldade e dá um critério muito eficaz para responder essa questão.

Teorema 7.3.4. *Seja E um espaço vetorial e $X = \{v_1, ..., v_m\} \subset E$. Se existe uma combinação linear nula*

$$\sum_{i=1}^{m} \lambda_i v_i = 0$$

tal que algum $\lambda_j \neq 0$ então X é LD.

Demonstração. Seja $j \in \{1, ..., m\}$ tal que $\lambda_j \neq 0$. Então, como $\sum_{i=1}^{m} \lambda_i v_i = 0$ podemos escrever

$$v_j = -\frac{1}{\lambda_j} \sum_{i=1, i\neq j}^{m} \lambda_i v_i = \sum_{i=1, i\neq j}^{m} \frac{-\lambda_i}{\lambda_j} v_i$$

Logo, v_j é combinação linear dos demais elementos de X, e portanto X é LD. □

Corollário 7.3.5. *Seja E um espaço vetorial e $X = \{v_1, ..., v_m\} \subset E$. Se X é LI e*

$$\sum_{i=1}^{m} \lambda_i v_i = 0$$

então $\lambda_j = 0$, para todo $j = 1, ..., m$.

Em termos práticos, o Corolário 7.3.5 estabelece que um conjunto X é LI se a única combinação linear dos elementos de X que dá zero for aquela em que todos os vetores de X são multiplicados por zero.

Exemplo 7.3.6. *Em $\mathbb{R}^d$, os vetores $e_1 = (1, 0, ..., 0)$, $e_2 = (0, 1, ..., 0)$,...,$e_d = (0, ..., 0, 1)$ formam um subconjunto LI. De fato, suponha que $\sum_{i=1}^{m} \lambda_i e_i = 0$. Então,*

$$(0, ..., 0) = (\lambda_1, 0, ..., 0) + (0, \lambda_2, ..., 0) + ... + (0, ..., 0, \lambda_d) = (\lambda_1, ..., \lambda_d),$$

e portanto $\lambda_i = 0$ para todo $i = 1, ..., m$. Pelo Corolário 7.3.5 concluímos que $\{e_1, ..., e_d\}$ é LI.

Consideremos agora a seguinte situação: seja $X = \{x_1, ..., x_m\} \subset E$ um subconjunto LI do espaço vetorial E. Seja $x \in \mathcal{S}(X)$. Então, por definição, existem números reais $\alpha_1, ..., \alpha_m$ tais que

$$x = \sum_{\ell=1}^{m} \alpha_\ell x_\ell.$$

Por exemplo, tomamos o espaço $E = \mathbb{R}^3$ e $X = \{(2, 0, 0), (0, \pi, 0)\}$. Claramente, $x = (1, \pi, 0)$ é combinação linear de $x_1 = (2, 0, 0)$ e $x_2 = (0, \pi, 0)$. Nesse caso, os coeficientes são $\alpha_1 = 0.5$ e

$\alpha_2 = 1$. Existe outra possibilidade de escolha de α_1 e α_2? Em outras palavras, os coeficientes da combinação linear de x_1 e x_2 que dá x são únicos?

Nesse caso podemos ver que a resposta é sim, porque qualquer outro coeficiente que multiplicar x_1 não vai dar 1 na primeira coordenada ao passo que $\alpha_2 x_2$ sempre tem a primeira coordenada igual a zero. Portanto, se $\alpha_1 \neq 0.5$, para nenhuma escolha de α_1, α_2 o vetor $\alpha_1 x_1 + \alpha_2 x_2$ tem 1 como primeira coordenada.

Considere agora, ainda em $\mathbb{R}^3$, $X = \{x_1 = (1, 2, 3), x_2 = (2, 4, 6)\}$. Nesse caso, o vetor $x = (4, 8, 12)$ pode ser escrito como

$$x = 2x_1 + x_2,$$

e nesse caso os coeficientes são $\alpha_1 = 2$ e $\alpha_2 = 1$, mas também como

$$x = 0.x_1 + 2x_2,$$

e nesse caso os coeficientes são $\alpha_1 = 0$ e $\alpha_2 = 2$. Nesse caso portanto, os coeficientes não são únicos.

O que diferencia um caso do outro? A resposta é que o primeiro conjunto é LI e o segundo é LD. De fato, como consequência do Corolário 7.3.5 temos outro corolário.

Corollário 7.3.7. *Se $X = \{x_1, ..., x_m\} \subset E$ é LI então para cada $x \in \mathcal{S}(X)$ e para cada $\ell = 1, ...m$ existe um único número real α_ℓ de forma que*

$$x = \sum_{\ell=1}^{m} \alpha_\ell x_\ell.$$

Demonstração. Suponhamos que existam α_ℓ e β_ℓ números reais tais que

$$x = \sum_{\ell=1}^{m} \alpha_\ell x_\ell = \sum_{\ell=1}^{m} \beta_\ell x_\ell.$$

Então, podemos escrever

$$\begin{aligned} 0 &= \sum_{\ell=1}^{m} \alpha_\ell x_\ell - \sum_{\ell=1}^{m} \beta_\ell x_\ell \\ &= \sum_{\ell=1}^{m} \alpha_\ell x_\ell - \beta_\ell x_\ell \\ &= \sum_{\ell=1}^{m} (\alpha_\ell - \beta_\ell) x_\ell. \end{aligned} \tag{7.1}$$

Ou seja, obtemos uma combinação linear nula dos elementos de X. Pelo Corolário 7.3.5, todos os coeficientes dessa combinação devem ser nulos e portanto $\alpha_\ell - \beta_\ell = 0$, para todo $\ell = 1, ..., m$. Isso prova que $\alpha_\ell = \beta_\ell$, para todo $\ell = 1, ..., m$ e completa a demonstração. □

7.4 Bases

Seja $X = \{x_1, ..., x_m\}$ um conjunto LD num espaço vetorial E. Discutimos acima (através de alguns exemplos específicos) que, intuitivamente, isso significa que existem em X vetores irrelevantes para construir o subespaço gerado $\mathcal{S}(X)$. Vamos tornar essa intuição precisa agora.

Lema 7.4.1 (Lema de Remoção dos Irrelevantes)**.** *Seja $X = \{y, x_1, ..., x_m\} \subset E$ tal que y é combinação linear de $\{x_1, ..., x_m\}$. Então $\mathcal{S}(\{x_1, ..., x_m\}) = \mathcal{S}(X)$.*

Demonstração. Existem m números reais $\alpha_1, ..., \alpha_m$ tais que

$$y = \sum_{j=1}^{m} \alpha_j x_j.$$

Essa igualdade significa, por definição que $x \in \mathcal{S}(Y)$. Nos resta então verificar que $\mathcal{S}(\{x_1, ..., x_m\}) = \mathcal{S}(X)$. Com efeito, como $\{x_1, ...x_m\} \subset X$, toda combinação linear dos elementos de $\{x_1, ..., x_m\}$ é uma combinação linear dos elementos de X e portanto

$$\mathcal{S}(\{x_1, ..., x_m\}) \subset \mathcal{S}(X).$$

Por outro lado, se $x \in \mathcal{S}(X)$ então existem $m + 1$ números reais $\beta, \beta_1, ...\beta_m$ tais que

$$x = \beta y + \sum_{j=1}^{m} \beta_j x_j.$$

Nessa o termo βy pode ser substituído por

$$\beta \sum_{j=1}^{m} \alpha_j x_j = \sum_{j=1}^{m} \beta \alpha_j x_j.$$

Pela distributividade segue que

$$x = \sum_{j=1}^{m} (\beta_j + \beta \alpha_j) x_j,$$

donde concluímos que $x \in \mathcal{S}(\{x_1, ..., x_m\})$. Provamos então que $\mathcal{S}(X) \subset \mathcal{S}(\{x_1, ..., x_m\})$, o que nos dá $\mathcal{S}(X) = \mathcal{S}(\{x_1, ..., x_m\})$. □

Vamos a seguir fazer uma segunda versão, ligeiramente mais técnica, do Lema de Remoção dos Irrelevantes. Nessa versão vamos mostrar como podemos "enfiar"um conjunto LI dentro de um conjunto gerador, retirando dele os elementos irrelevantes.

Lema 7.4.2 (Lema de Remoção dos Irrelevantes-Versão Ampliada)**.** *Sejam $Y = \{y_1, ..., y_k\}$ e $X = \{x_1, ..., x_m\}$ subconjuntos de E e $y \in E$ tais que $Y \cup \{y\}$ é LI e $X \cup Y$ é gerador. Então exite $j \in \{1, ..., m\}$ tal que se*

$$Z = \{y\} \cup Y \cup X \setminus \{x_j\}$$

então Z é gerador

Demonstração. Como $X \cup Y$ é gerador, o conjunto $\{y\} \cup X \cup Y$ é gerador e LD. Portanto existe uma combinação linear nula de seus elementos,

$$\alpha y + \alpha_1 y_1 + ... + \alpha_k y_k + \beta_1 x_1 + ... + \beta_m x_m = 0, \tag{7.2}$$

na qual nem todos os coeficientes são nulos.

Afirmo que existe pelo menos um $j \in \{1, ..., m\}$ tal que $\beta_j \neq 0$.

Com efeito, se não fosse assim teríamos $\beta_1 = ... = \beta_m = 0$ e portanto

$$\alpha y + \alpha_1 y_1 + ... + \alpha_k y_k = 0.$$

Como $\{y\} \cup Y$ é LI, deveríamos ter $\alpha = \alpha_1 = ...\alpha_k = 0$, e assim teríamos todos os coeficientes na combinação linear (7.2) nulos, o que seria absurdo.

Tome agora qualquer elemento $j \in \{1, ..., m\}$ para o qual $\beta_j \neq 0$. Então podemos escrever

$$x_j = \frac{\alpha}{\beta_j} y + \frac{\alpha_1}{\beta_j} y_1 + ... + \frac{\alpha_k}{\beta_j} y_k + \sum_{i=1, i \neq j}^{m} \frac{\beta_i}{\beta_j} x_i.$$

Ou seja, x_j é combinação linear dos elementos de Z. Aplicando a primeira versão do lema de remoção dos irrelevantes segue que retirando-se x_j do conjunto $\{y\} \cup X \cup Y$ temos ainda um conjunto gerador. Isso prova o lema. □

Nesse momento vamos atingir uma etapa crucial nesse curso. Vamos ver uma demonstração astuta e linda exclusivamente de *álgebra linear*. O que iremos provar é bastante razoável: um subconjunto LI nunca pode ter mais elementos do que um gerador. Mais imagine isso no espaço dos polinômios de grau menor do que 478!!! É para esse tipo de dificuldade que precisamos de demonstrações. Melhor ainda quando elas são lindas.

Lema 7.4.3 (Gerador domina LI)**.** *Se $x_1, ..., x_m$ geram o espaço vetorial E e $y_1, ..., y_n$ é LI então $n \leq m$.*

Demonstração. Seja $X = \{x_1, ..., x_m\}$. Como X é gerador, adicionando-se qualquer vetor de E à X obtemos um conjunto LD. Em particular, $\{y_1, x_1, ..., x_m\}$ é LD. Como $Y = \{y_1, ..., y_n\}$ é LI devemos ter $y_1 \neq 0$. O lema de remoção dos irrelevantes então nos permite retirar um dos $x_i's$ do conjunto $\{y_1, x_1, ..., x_m\}$ e ainda ter um gerador. Seja $\mathcal{V}_1$ o conjunto obtido após essa remoção.

Adicionamos agora y_2 à $\mathcal{V}_1$. Mais uma vez, como $\mathcal{V}_1$ é gerador obtemos o conjunto $\{u_2\} \cup \mathcal{V}_1$ que deve ser LD. Como $\{y_1, y_2\}$ é LI, o lema de remoção dos irrelevantes mais uma vez nos perite retirar um dos $x_i's$ do conjunto $\{u_2\} \cup \mathcal{V}_1$ e ainda ter um gerador. Seja $\mathcal{V}_2$ o conjunto obtido após essa remoção.

Prosseguimos assim, acrescentando y's e retirando x's. Ou seja, na etapa j, teremos um conjunto gerador $\mathcal{V}_j$ com j elementos Y, e $m - j$ elementos de X.

Note que nunca poderemos retirar todos os x's antes de colocar todos os y's porque isso implicaria que Y seria LD. Então podemos colocar todos os y's e talvez ainda sobre um x. Portanto, tem mais x's do que y's. Ou seja, $m \geq n$, como queria demonstrar. □

Definição 7.4.4. *Seja E um espaço vetorial. Uma* base *de E é um subconjunto gerador e LI.*

Exemplo 7.4.5. *O conjunto $\{e_1, ..., e_d\}$ é uma base do espaço vetorial $\mathbb{R}^d$, chamada a* canônica *de $\mathbb{R}^d$.*

7.4.1 Espaços vetoriais de dimensão finita

Um espaço vetorial é dito de *dimensão finita* quando possui uma base com um número finito de elementos. *Neste livro, iremos estudar apenas espaços vetoriais de dimensão finita.*

Ou seja, a menos de menção explícita em contrário, em toda expressão "seja E um espaço vetorial", fica subentendido que E possui dimensão finita.

Diante do conceito de base de um espaço vetorial, certas perguntas saltam aos olhos: é possível que um mesmo espaço vetorial possua duas bases com quantidades distintas de elementos? É possível encontrar um conjunto LI com mais do que d vetores em $\mathbb{R}^d$? Existe um conjunto de geradores de $\mathbb{R}^d$ com menos do que d vetores?

A seguir, nos dedicaremos a responder cada uma dessas perguntas. Primeiro, vamos ver que quaisquer duas bases num espaço vetorial tem o mesmo número de elementos. A demonstração desse fato é um corolário do Lema 7.4.3

Corollário 7.4.6. *Se o espaço vetorial E admite uma base $\mathcal{B} = \{b_1, ..., b_d\}$ e se $X = \{v_1, ..., v_m\}$ é também uma base então $m = d$.*

Demonstração. Como $\mathcal{B}$ é gerador e X é LI, temos que $m \leq d$. Como X é gerador e $\mathcal{B}$ é LI, temos que $d \leq m$. □

Exemplo 7.4.7. *O espaço $\mathbb{R}^3$ é gerado por 3 vetores, os vetores $\{e_1, e_2, e_3\}$ da base canônica. Além disso, geometricamente, podemos ver que um conjunto LI com três vetores é gerador (veja a Figura 7.2). Logo, se adicionarmos um quarto vetor, ele já será combinação linear dos demais. Isso mostra que qualquer conjunto com 4 vetores em $\mathbb{R}^3$ é LD.*

A *dimensão* de um espaço vetorial E, que denotamos por $\dim E$, é o número de elementos de uma base qualquer de E. Assim, pelo Exemplo 7.4.5 concluímos que

$$\dim \mathbb{R}^d = d,$$

para todo $d \geq 1$.

A presença de bases em espaços vetoriais faz com que possamos trabalhar (do ponto de vista computacional) num espaço vetorial de dimensão d da mesma forma como trabalhamos no espaço euclideano $\mathbb{R}^d$.

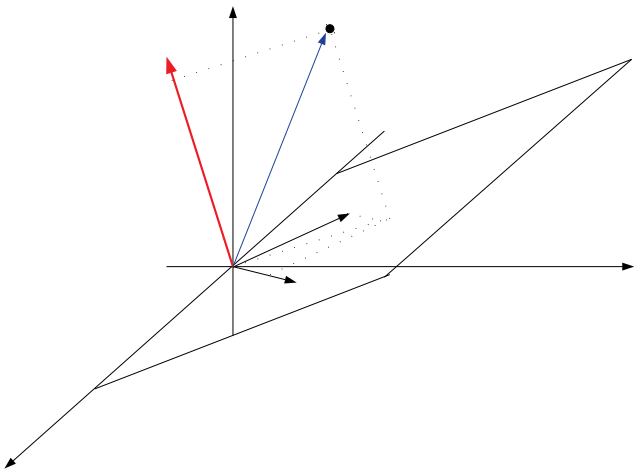

Figura 7.2: Com dois vetores geramos um plano. Adicionando-se um vetor de fora desse plano geramos todo o $\mathbb{R}^3$.

Com efeito, seja E um espaço vetorial, $d = \dim E$. Seja $\mathcal{B} = \{b_1, ..., b_d\}$ uma base de E. Então, dado $x \in E$, existem números reais $x_1, ..., x_d$ tais que

$$x = \sum_{i=1}^{d} x_i b_i.$$

Assim, o ponto $x \in E$ pode ser identificado com o ponto $(x_1, ..., x_d) = \sum_{i=1}^{d} x_i e_i \in \mathbb{R}^d$. Os números reais $(x_1, ..., x_d)$ são chamados *coordenadas* do vetor x relativamente à base $\mathcal{B} = \{b_1, ..., b_d\}$.

Teorema 7.4.8. *Seja E um espaço vetorial. Seja $X = \{x_1, ..., x_m\} \subset E$ um subconjunto LI, que não é gerador. Então existe $Y = \{y_1, ..., y_n\} \subset E$ tal que $\mathcal{B} = \{x_1, ..., x_m, y_1, ..., y_n\}$ é uma base de E. Em outras palavras, num espaço vetorial todo conjunto LI pode ser completado de modo a se tornar uma base.*

Demonstração. Seja $F = \mathcal{S}(X)$. Como o complementar $E - F$ é não vazio, existe um vetor $y_1 \in E - F$. Como $y \notin \mathcal{S}(X)$, o conjunto $X_1 = X \cup \{y_1\}$ é LI. Se $\mathcal{S}(X_1) = E$, acabou. Caso contrário, existe $y_2 \in E - \mathcal{S}(X_1)$, e temos, como antes, que $X_2 = X_1 \cup \{y_2\}$ é LI. Prosseguindo dessa maneira, obtemos conjuntos LI $X_n \subset X_{n+1}$ cada vez maiores. Como E tem dimensão finita, um desses conjuntos será gerador e portanto será a base procurada. □

Com o mesmo tipo de raciocínio do teorema anterior, podemos demonstrar o seguinte

Teorema 7.4.9. *Todo conjunto de geradores contém uma base. Se $F \subset E$ é um subespaço então* $\dim F \leq \dim E$. *Se* $\dim F = \dim E$ *então* $E = F$.

Demonstração. Suponha por contradição que $E \setminus F \neq \emptyset$. Seja $\mathcal{B} = \{b_1, ..., b_d\}$ uma base de F. Como $E \setminus F \neq \emptyset$, existe $b \in E \setminus F$. Como $F = \mathcal{S}(\mathcal{B})$, temos que b não é combinação linear dos elementos de $\mathcal{S}(\mathcal{B})$. Afirmo que o conjunto $\{b_1, .., b_d, b\}$ é LI. Com efeito, se não fosse haveriam números reais $\alpha_1, .., \alpha_d, \alpha$, nem todos nulos, tais que

$$\alpha_1 b_1 + ... + \alpha_d b_d + \alpha b = 0.$$

Devemos ter $\alpha \neq 0$, pois do contrario teríamos

$$\sum_{\ell=1}^{d} \alpha_\ell b_\ell = 0,$$

e como $\mathcal{B}$ é LI, concluiríamos que os números $\alpha_1, .., \alpha_d, \alpha$ são todos nulos. No entanto, se $\alpha neq 0$ podemos escrever

$$b = \sum_{\ell=1}^{d} -\frac{\alpha_\ell}{\alpha} b_\ell \in \mathcal{S}(\mathcal{B}),$$

contradição. Como $d = \dim F = \dim E$, por hipótese, usando o Lema 7.4.3, temos que

$$d + 1 \leq d \implies 1 \leq 0,$$

um absurdo. Isso completa a demonstração. □

Assim, a estrutura de um espaço vetorial pode ser completamente entendida através do conceito de base: todo espaço vetorial (aí incluso, obviamente, os subespaços de um espaço vetorial) é o conjunto de combinações lineares dos elementos da base, a qual sempre existe. Além disso, um vetor sempre pode ser identificado com suas *coordenadas* com relação à base dada.

Em todo espaço vetorial, sempre existem infinitas bases. A escolha de uma base depende fundamentalmente do problema de interesse. Veremos nos próximos capítulos situações onde a escolha de uma base adequada pode ser decisiva na solução de um dado problema.

Capítulo 8

Bases ortonormais

Seja E um espaço vetorial de dimensão finita d, munido de um produto interno $\langle ., . \rangle$. Uma base $\mathcal{B} = \{b_1, ..., b_d\}$ de E é dita *ortonormal* se satisfaz

$$\langle b_i, b_j \rangle = \begin{cases} 1, & \text{se } i = j \\ 0, & \text{caso contrário} \end{cases} .$$

Por exemplo, a base canônica $\mathcal{E} = \{e_1, ..., e_d\}$ de $\mathbb{R}^d$ munido do produto interno euclideano é um exemplo de uma base ortonormal.

As bases ortonormais permitem dotar qualquer espaço vetorial munido de um produto interno da mesma estrutura algébrica e geométrica dos espaços euclideanos. Além, mesmo no contexto específico dos espaços euclideanos sua importância conceitual não deve ser negligenciada.

Por isso, neste capítulo vamos aprender um algorítimo para construir bases ortonormais a partir de qualquer base e vamos ver uma aplicação geométrica desse algorítimo.

8.1 O algorítimo de Gram-Schmidt

Seja E um espaço vetorial de dimensão finita d, munido de um produto interno $\langle ., . \rangle$. O leitor menos interessado em generalidades matemáticas pode sem problemas ler tudo o que se segue

assumindo que $E = \mathbb{R}^d$ e que o produto interno em tela é o produto interno euclideano usual.

Seja $F \subset E$ um subespaço de dimensão n e seja $\{v^1, ..., v^n\} \subset F$ uma base qualquer de F.

Teorema 8.1.1 (Gram-Schmidt)**.** *Existe* $\{w^1, ..., w^n\} \subset F$ *tal que* $\langle w^i, w^j \rangle = 0$ *para todo* $i \neq j$ *e, além disso, para cada* $\ell = 1, ..., n$*,*

$$\mathcal{S}\left(\{v^1, ..., v^\ell\}\right) = \mathcal{S}\left(\{w^1, ..., w^\ell\}\right).$$

Demonstração. Defina $w^1 = v^1$ e vamos definir w^2 pondo

$$w^2 \stackrel{\text{def}}{=} v^2 - \frac{\langle v^2, w^1 \rangle}{\langle w^1, w^1 \rangle} w^1.$$

Então, claramente $w^2 \in \mathcal{S}\left(\{v^2, v^1\}\right)$ e

$$\langle w^2, w^1 \rangle = \langle v^2 - \frac{\langle v^2, w^1 \rangle}{\langle w^1, w^1 \rangle} w^1, w^1 \rangle = \langle v^2, w^1 \rangle - \frac{\langle v^2, w^1 \rangle}{\langle w^1, w^1 \rangle} \langle w^1, w^1 \rangle = 0.$$

Vamos então mostrar como de uma etapa da construção se passa para a etapa seguinte.

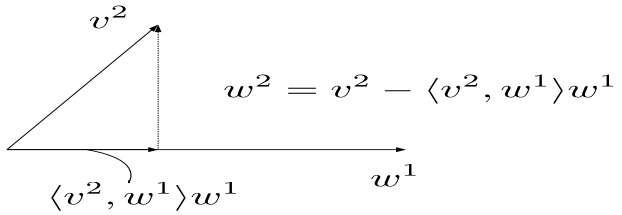

Figura 8.1: O primeiro passo do algorítimo de Gram-Schmidt

Suponha por indução que os k primeiros vetores $w^1, ..., w^k$ tenham sido construídos. Evidentemente, estamos pensando nesse raciocínio que $k < \ell$. Defina

$$w^{k+1} \stackrel{\text{def}}{=} v^{k+1} - \sum_{j=1}^{k} \frac{\langle v^{k+1}, w^j \rangle}{\langle w^j, w^j \rangle} w^j.$$

Observe que $w^{k+1} \in \mathcal{S}(\{w^1, ..., w^k, v^{k+1}\})$. Como estamos supondo, seguindo o princípio da indução finita, que as primeiras k etapas foram realizadas com sucesso, sabemos que

$\mathcal{S}(\{v^1, ..., v^k\}) = \mathcal{S}(\{w^1, ..., w^k\})$. Logo, $\mathcal{S}(\{v^1, ..., v^k, v^{k+1}\}) = \mathcal{S}(\{w^1, ..., w^k, v^{k+1}\})$. Portanto, $w^{k+1} \in \mathcal{S}(\{v^1, ..., v^k, v^{k+1}\})$ e concluímos assim que

$$\mathcal{S}(\{v^1, ..., v^k, v^{k+1}\}) = \mathcal{S}(\{w^1, ..., w^k, w^{k+1}\}).$$

Vamos agora verificar que w^{k+1} é ortogonal a todos os vetores já construídos. Com efeito, pela linearidade do produto interno temos que para todo $i = 1, ..., k$,

$$\langle w^{k+1}, w^i\rangle = \langle v^{k+1} - \sum_{j=1}^{k} \frac{\langle v^{k+1}, w^j\rangle}{\langle w^j, w^j\rangle} w^j, w^i = \langle v^{k+1}, w^i\rangle - \sum_{j=1}^{k} \frac{\langle v^{k+1}, w^j\rangle}{\langle w^j, w^j\rangle} \langle w^j, w^i\rangle.$$

Como as etapas anteriores foram executadas com sucesso, temos que para cada $j \neq i$, $\langle w^j, w^i\rangle = 0$. Portanto, a única parcela não-nula no somatório acima vai ser aquela correspondente a $j = i$. Segue que

$$\langle w^{k+1}, w^i\rangle = \langle v^{k+1}, w^i\rangle - \langle v^{k+1}, w^i\rangle \frac{\langle w^i, w^i\rangle}{\langle w^i, w^i\rangle} = 0.$$

Isso prova, por indução, que todas as etapas serão construídas com sucesso e conclui a demonstração. □

8.2 O complemento ortogonal

Vamos agora apresentar uma aplicação conceitual (mas com forte apelo geométrico) do Teorema 8.1.1. Seja $F \subset E$ um subespaço de dimensão n.

Definição 8.2.1. *O* complemento ortogonal *de F é o conjunto*

$$F^\perp \stackrel{\text{def}}{=} \{y \in E; \langle y, x\rangle = 0, \forall\, x \in F\}.$$

Ou seja, o complemento ortogonal de F é conjunto formado pelos elementos de E que são ortogonais a todos os vetores em F.

Lema 8.2.2. *$F^\perp$ é um subespaço*

Demonstração. Como 0 é ortogonal a todos os vetores de E, temos em particular que $0 \in F^\perp$. Suponha agora que $x, y \in F^\perp$ e $\lambda \in \mathbb{R}$. Vamos provar que o vetor $x + \lambda y$ pertence a $F^\perp$. Isso

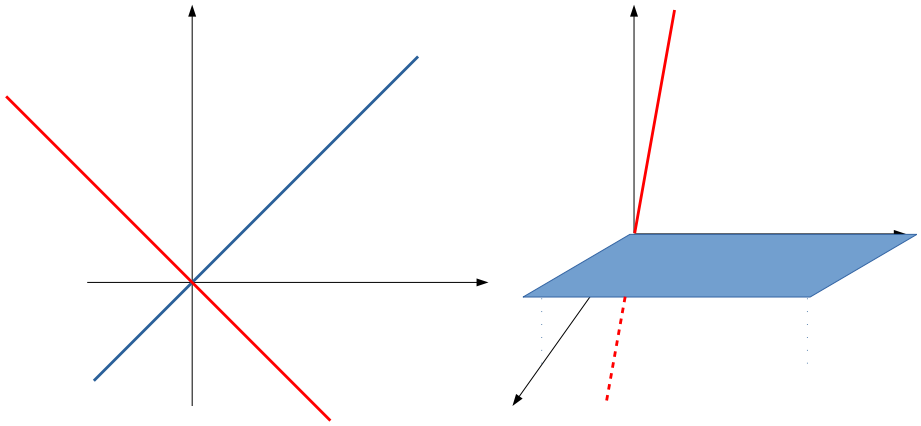

Figura 8.2: Nas duas figuras, o subespaço vermelho é o complemento ortogonal do subespaço azul.

significa provar que $x + \lambda y$ é ortogonal a todos os vetores em F. Então, seja $z \in F$ qualquer. Temos, pela linearidade do produto interno que

$$\langle z, x + \lambda y \rangle = \langle z, x \rangle + \lambda \langle z, y \rangle.$$

Como $x, y \in F^\perp$, temos $\langle x, z \rangle = \langle y, z \rangle = 0$, logo $\langle z, x + \lambda y \rangle = 0$. Portanto $x + \lambda y \in F^\perp$ para todo $\lambda \in \mathbb{R}$. □

Lema 8.2.3. $F^\perp \cap F = \{0\}$

Demonstração. Suponha por contradição que exista $y \in F^\perp \cap F \setminus \{0\}$. Como $y \in F^\perp$, y é ortogonal a todos os vetores de F. Como $y \in F$, segue que y é ortogonal a si próprio, ou seja

$$\|y\|^2 = \langle y, y \rangle = 0.$$

Mas isto implica que $y = 0$, absurdo. □

Lema 8.2.4. $\dim F^\perp \leq d - n$

Demonstração. Vamos denotar $\ell = d - n$. Sejam $\{v^1, ..., v^n\} \subset F$ uma base de F e seja $\{u^1, ..., u^k\}$ uma base de $F^\perp$. Como $F^\perp \cap F = \{0\}$, nenhum vetor u^j é combinação linear dos vetores $v^1, ..., v^n$. Portanto, o conjunto $\{v^1, ..., v^n, u^1, ..., u^k\} \subset E$ é LI. Pelo Lema 7.4.3 temos então que $n + k \leq d \implies k \leq d - n$. □

Observe com atenção a figura 8.2. No lado, esquerdo, F (a reta azul) e $F^{\perp}$ a reta vermelha possuem ambos dimensão 1, portanto a soma das suas dimensões é a dimensão do espaço ambiente. NA figura do lado direito, F é o plano azul, que tem dimensão 2 e $F^{\perp}$ é a reta vermelha que têm dimensão 1, portanto a soma de suas dimensões é igual a dimensão do espaço todo. Vamos agora aplicar o algorítimo de Gram-Schmidt para ver que nada disso é mera coincidência.

Teorema 8.2.5. $\dim F + \dim F^{\perp} = d$

Demonstração. Seja $\{v^1, ..., v^n\}$ uma base de F. Complete essa base (usando o Teorema 7.4.8) com os vetores $u^1, ..., u^{\ell}$, onde $\ell = d - n$, de modo que o conjunto

$$v^1, ..., v^n, u^1, ..., u^{\ell}$$

seja uma base de E. Aplicando o Teorema 8.1.1 obtemos uma nova base

$$w^1, ..., w^n, w^{n+1}, ..., w^{n+\ell}$$

de modo que

$$\mathcal{S}\left(\{w^1, ..., w^n\}\right) = \mathcal{S}\left(\{v^1, ..., v^}\right) = F,$$

e como os vetores w^j são todos ortogonais entre si, vemos que $\{w^{n+1}, ..., w^{n+\ell}\}$ é um conjunto LI contido em $F^{\perp}$. Isso implica que $\dim F^{\perp} \geq \ell$. Pelo Lema 8.2.4 segue que $\dim F^{\perp} = \ell$. Como $\ell + n = d$, a prova está terminada. □

Capítulo 9

Funções lineares e suas matrizes

A estrutura algébrica fundamental de um espaço vetorial está encapsulada no conceito de bases, que vimos anteriormente. Todo espaço vetorial de dimensão finita é construído com uma quantidade finita de informação: todo vetor pode ser escrito como combinação linear dos elementos da base.

Neste capítulo vamos olhar essa estrutura de uma outra perspectiva, mais geométrica. Vamos estudar funções entre espaços vetoriais. Contudo, o que há de geometria nisso?

Ora, uma função $f : X \to Y$ entre dois conjuntos pode ser olhada como uma transformação: uma entidade que transforma elementos do conjunto X em elementos do conjunto Y. Se X e Y forem dotados da estrutura de espaço vetorial então, uma função $f : X \to Y$ transforma pontos do espaço X em pontos do espaço Y. Agora, imagine que $X = Y = \mathbb{R}^3$ só para fixar ideias. Como uma reta em $\mathbb{R}^3$ é transformada por f? Como um plano em $\mathbb{R}^3$ é transformado por f? Como uma esfera em $\mathbb{R}^3$ é transformado por f? O sabor dessas perguntas é evidentemente geométrico. No entanto, a estrutura algébrica de um espaço vetorial permite que essas mesmas perguntas possam ser formuladas em qualquer espaço vetorial de dimensão finita!

Com efeito, uma reta num espaço vetorial é o subespaço gerado por um vetor. Um plano é o subespaço gerado por dois vetores. Uma esfera é o conjunto dos vetores cujo quadrado das coordenadas com respeito a uma base ortonormal somam 1.[1]

[1]Ou seja, se E é um espaço vetorial de dimensão d e $\mathcal{B} = \{b_1, ..., b_d\}$ é uma base ortonormal então uma

Então, em resumo, vamos estudar esse tipo pergunta geométrica sobre funções (ou transformações) entre espaços vetoriais com o intuito de mergulhar mais profundamente dentro da estrutura algébrica deles. Por essa razão, vamos restringir esse estudo a um tipo muito especial de transformações: justamente aquelas que respeitam a estrutura algébrica, as transformações lineares.

Ao leitor que só está aqui por que assim manda a graduação em economia, ou engenharia, eu sei que isso pode parecer metafísica esotérica de matemáticos que já têm a vida ganha e nada com que se preocupar. Pois justamente, as aplicações desse estudo que veremos visam te mostrar, caro leitor, que essa metafísica esotérica é coisa muito séria, que molda a realidade a nossa volta, e que o entendimento dela vai ter um impacto profundo na sua carreira (e portanto na sua conta bancária).

9.1 Funções Lineares

Lembre que, se X e Y são conjuntos uma função entre X e Y é uma lei que associa a cada elemento $x \in X$ um elemento $f(x) \in Y$, chamado imagem de x pela função $f : X \to Y$.

Definição 9.1.1. *Sejam E e F espaços vetoriais. Uma função $A : E \to F$ entre E e F é dita* linear *se satisfaz as seguintes propriedades*

1. *$A(x + y) = A(x) + A(y)$, para todos $x, y \in E$;*

2. *$A(\lambda x) = \lambda A(x)$, para todos $\lambda \in \mathbb{R}$ e $x \in E$.*

As funções entre espaços vetoriais que são dotadas do superpoder da linearidade são, em muitos lugares literatura afora, também chamadas de *transformações lineares*, um nome que enfatiza o sabor geométrico desse ingrediente fundamental da álgebra linear. Embora eu prefira evitar essa nomenclatura, pois ela obscurece o fato de que uma transformação linear nada mais é do que uma função dotada de uma propriedade especial, é possível que eu chame funções lineares aqui e acolá de transformações lineares.

---

esfera em E é o conjunto $\mathbb{S}^{d-1} \stackrel{\text{def}}{=} \{x \in E; x = \sum_{i=1}^{d} x_i b_i; \sum_{i=1}^{d} x_i^2 = 1\}$.

Exemplo 9.1.2. *1. Lembre que $\mathbb{R}$ é um espaço vetorial de dimensão 1. A função $A : \mathbb{R} \to \mathbb{R}$, definida por $A(x) = 2x$ é linear. Com efeito,*

$$A(x + y) = 2(x + y) = 2x + 2y = A(x) + A(y),$$

e

$$A(\lambda x) = 2\lambda x = \lambda 2x = \lambda A(x).$$

2. Considere $f : \mathbb{R}^3 \to \mathbb{R}$ definida por

$$f(x, y, z) = x + y + z.$$

Então, f é uma transformação linear entre os espaços vetoriais $\mathbb{R}^3$ e $\mathbb{R}$. Com efeito, se (x_1, y_1, z_1) e (x_2, y_2, z_2) são vetores em $\mathbb{R}^3$ temos que

$$\begin{aligned} f((x_1, y_1, z_1) + (x_2, y_2, z_2)) &= f(x_1 + x_2, y_1 + y_2, z_1 + z_2) && (9.1)\\ &= x_1 + x_2 + y_1 + y_2 + z_1 + z_2 && (9.2)\\ &= x_1 + y_1 + z_1 + x_2 + y_2 + z_2 && (9.3)\\ &= f(x_1, y_1, z_1) + f(x_2, y_2, z_2). && (9.4) \end{aligned}$$

Analogamente, se $(x, y, z) \in \mathbb{R}^3$ e $\lambda \in \mathbb{R}$ então

$$f(\lambda x, \lambda y, \lambda z) = \lambda x + \lambda y + \lambda z = \lambda(x + y + z) = \lambda f(x, y, z).$$

3. Sejam E e X um espaços vetoriais quaisquer. Seja $\zeta : E \to X$ dada por $\zeta(x) = 0$. Então, ζ é uma transformação linear.

4. Seja E um espaço vetorial e $\mathrm{I} : E \to E$ dada por $\mathrm{I}(v) = v$. Então ι é uma transformação linear, chamada a identidade *do espaço E.*

Uma transformação linear é uma função que preserva a estrutura linear de um espaço vetorial. Assim, como veremos logo a seguir, sempre transforma subespaços em subespaços. De fato, é fácil ver que uma transformação linear $A : E \to F$ sempre transforma o vetor nulo de E no vetor nulo F.

Lema 9.1.3. *Sejam E e F espaços vetoriais e $A : E \to F$ uma transformação linear. Então $A(0) = 0$.*

Demonstração. Com efeito, $A(0) = A(0+0) = A(0) + A(0)$, portanto $A(0) = A(0) - A(0) = 0$. □

9.1.1 Matrizes como transformações lineares

Uma matriz $A_{m\times n} = [a_{ij}]$ é uma tabela de números com m linhas e n colunas, onde o número que fica na interseção da linha i com a coluna j é o número a_{ij}.

Exemplo 9.1.4. *Exemplos de matrizes:*

1. *Uma matriz* 2×2 $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$
2. *Uma matriz* 3×2 $A = \begin{bmatrix} 3 & 7 & 4 \\ \pi & e^2 & 0 \end{bmatrix}$.

Dada uma matriz, $A = [a_{ij}]_{m\times n}$, podemos decompô-la em seus vetores-linha e vetores-coluna. Observe que cada coluna A possui m entradas, logo pode ser vista como um vetor de $\mathbb{R}^m$, ao passo que cada linha, possuindo n entradas, pode ser vista como um vetor de $\mathbb{R}^n$.

Definição 9.1.5 (Vetor-linha e vetor-coluna)**.** *Seja $A = [a_{ij}]_{m\times n}$ uma matriz com m linhas e n colunas. Para cada $j \in \{1, ..., n\}$, o vetor*

$$c_j \stackrel{\text{def}}{=} (a_{1j}, ..., a_{mj}) \in \mathbb{R}^m.$$

é chamado o j-ésimo vetor coluna *de A. Para cada $i \in \{1, ..., m\}$ o vetor*

$$\ell_i \stackrel{\text{def}}{=} (a_{i1}, ..., a_{in}) \in \mathbb{R}^n$$

é chamado o i–ésimo vetor linha *de A.*

Abuso de nomeclatura– É importante o leitor ter em mente, daqui para frente, que, por motivo de fluidez na comunicação, na maioria das vezes que nos referirmos a um vetor linha de uma matriz A vamos falar somente uma linha de A. Analogamente para os vetores coluna.

Toda matriz $A_{m\times n}$ pode ser encarada como uma função linear $A : \mathbb{R}^n \to \mathbb{R}^m$. Vamos explicar este procedimento em detalhes. Dado um vetor $x = (x_1, ..., x_n) \in \mathbb{R}^n$, podemos escrevê-lo como uma matriz $n \times 1$:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Observe que a matriz $A_{m\times n}$ possui n de elementos em cada linha, o mesmo número de entradas do vetor $x = (x_1, ..., x_n) \in \mathbb{R}^n$. Assim, para cada linha i da matriz podemos multiplicar o termo a_{ij} pela coordenada x_j. A soma dessas n multiplicações será a i-ésima coordenada do vetor imagem $A(x) \in \mathbb{R}^m$.

$$\begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \dots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^n a_{1j}x_j \\ \vdots \\ \sum_{j=1}^n a_{mj}x_j \end{bmatrix}$$

Observe que a i-ésima coordenada do vetor imagem $A(x)$ é o produto interno $\langle \ell_i, x\rangle$ entre a i-ésima linha da matriz A com o vetor x. Logo, podemos escrever

$$A(x) = \begin{bmatrix} \langle \ell_1, x\rangle \\ \vdots \\ \langle \ell_m, x\rangle \end{bmatrix}.$$

A função $A : \mathbb{R}^n \to \mathbb{R}^m$ assim definida é uma transformação linear, pois se $x, y \in \mathbb{R}^n$ então, pela linearidade do produto interno euclideano

$$A(x+y) = \begin{bmatrix} \langle \ell_1, x+y\rangle \\ \vdots \\ \langle \ell_m, x+y\rangle \end{bmatrix} = \begin{bmatrix} \langle \ell_1, x\rangle \\ \vdots \\ \langle \ell_m, x\rangle \end{bmatrix} + \begin{bmatrix} \langle \ell_1, y\rangle \\ \vdots \\ \langle \ell_m, y\rangle \end{bmatrix} = A(x) + A(y),$$

e, analogamente, qualquer que seja $\lambda \in \mathbb{R}$, teremos

$$A(\lambda x) = \begin{bmatrix} \langle \ell_1, \lambda x\rangle \\ \vdots \\ \langle \ell_m, \lambda x\rangle \end{bmatrix} = \begin{bmatrix} \lambda\langle \ell_1, x\rangle \\ \vdots \\ \lambda\langle \ell_m, x\rangle \end{bmatrix} = \lambda A(x).$$

Exemplo 9.1.6. *1. Seja* $A = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$, *uma matriz quadrada* 2×2. *Então, a transformação linear* $A : \mathbb{R}^2 \to \mathbb{R}^2$ *induzida por* A *é a função que leva um vetor* $v = (x, y) \in \mathbb{R}^2$ *no vetor* $A(v) = (2x + y, x + y)$, *pois*

$$\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2x + y \\ x + y \end{bmatrix}.$$

Observe que $A(0, 0) = (0, 0)$ *(já sabíamos que tinha que ser assim devido ao Lema9.1.3). Além disso,* $A(1, 2) = (4, 3)$ *e* $A(2, 4) = (8, 6)$. *Esta última igualdade apenas ilustra o fato, também já esperado, que* $A(2(1, 2)) = 2A(1, 2)$. *Mais geralmente, a reta*

$$r = \{(\lambda, 2\lambda); \lambda \in \mathbb{R}\}$$

é transformada por A *na reta*

$$A(r) = \{(4\lambda, 3\lambda); \lambda \in \mathbb{R}\}.$$

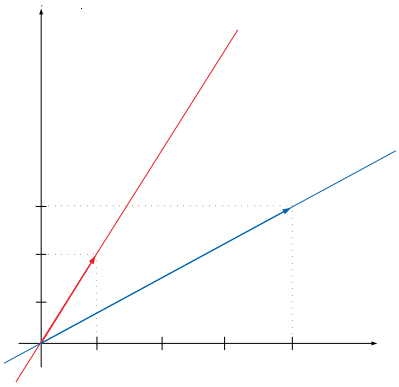

Figura 9.1: A matriz A transforma a reta vermelha, gerada pelo vetor $(1, 2)$, na reta azul, gerada pelo vetor $(4, 3)$.

2. Uma matriz pode colapsar uma reta na origem. Isso é fácil de ver na matriz $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, *pois o vetor* $e_1 = (1, 0)$ *é transformado no vetor* $A(1, 0) = (0, 0)$, *mas também*

ocorre com a matriz $A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$. *Com efeito, o vetor* $(2, -1)$ *é transformado no vetor*

$$A(2,-1) = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = (2-2, 4-4) = (0,0).$$

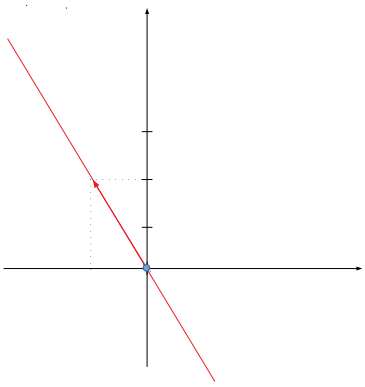

Figura 9.2: A matriz A colapsa a reta vermelha, gerada pelo vetor $(2, -1)$, no ponto $(0, 0)$.

Como descobrir de antemão qual vetor será ou não colapsado pela matriz? No caso específico deste exemplo, considere os vetores $e_1 = (1, 0)$ *e* $e_2 = (0, 1)$ *da base canônica de* $\mathbb{R}^2$. *Observe que* $A(e_1) = (1, 2)$ *e* $A(e_2) = (2, 4) = 2A(e_1)$. *Logo,*

$$\begin{aligned} 0 &= A(e_2) - 2A(e_1) = A(e_2) - A(2e_1) && (9.5) \\ &= A(e_2 - 2e_1) = A((0,1) - (2,0)) = A(-2,1) && (9.6) \\ &= -A(2,-1), && (9.7) \end{aligned}$$

e portanto $A(2, -1) = 0$. *De um ponto de vista mais conceitual, podemos esclarecer esse exemplo da seguinte forma: a base o vetor* e_2 *é transformado num múltiplo da imagem do vetor* e_1*; ou seja: o conjunto LI* $\{e_1, e_2\}$ *é transformado num conjunto LD* $\{A(e_1), A(e_2)\}$.

3. *Considere a transformação linear* $f : \mathbb{R}^3 \to \mathbb{R}$, *do exemplo 9.1.2. Note que* f *pode ser vista a transformação linear induzida pela matriz de uma única linha* $\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$, *pois*

$$\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = x + y + z.$$

Em particular, $f(e_1) = f(e_2) = f(e_3) = 1$. *Por outro lado,* $f(1,1,-2) = 0$ *e* $f(1,2,-3) = 0$. *Observe que os vetores* $(1,1,-2)$ *e* $(1,2,-3)$ *não são colineares, e portanto geram um plano em* $\mathbb{R}^3$. *Esse plano é o subespaço gerado por* $(1,1,-2)$ *e* $(1,2,-3)$,

$$P = \{\alpha(1,1,-2) + \beta(1,2,-3); \alpha, \beta \in \mathbb{R}\},$$

que é o mesmo que o conjunto de todas as combinações lineares possíveis entre $(1,1,-2)$ *e* $(1,2,-3)$. *Como* f *é linear, para todo* $v \in P$, *teremos*

$$\begin{aligned} f(v) &= f(\alpha(1,1,-2) + \beta(1,2,-3)) \\ &= f(\alpha(1,1,-2)) + f(\beta(1,2,-3)) \\ &= \alpha f(1,1,-2) + \beta f(1,2,-3) \\ &= \alpha 0 + \beta 0 = 0. \end{aligned} \tag{9.8}$$

Portanto, f *colapsa o plano* P *em zero.*

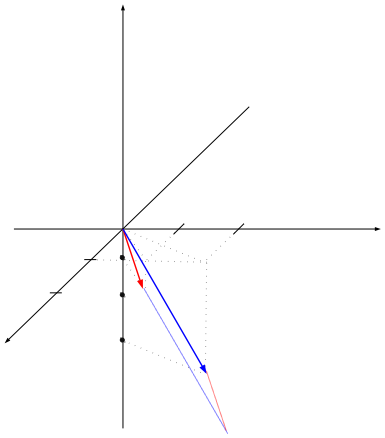

Figura 9.3: A transformação f colapsa o plano P, gerado pelos vetores azul e vermelho, em zero.

Exemplo 9.1.7. *Nos dois exemplos a seguir iremos analisar transformações lineares definidas no espaço* $\mathbb{R}^3$. *Lembre que a base canônica do* $\mathbb{R}^3$ *é formada pelos vetores* $e_1 = (1,0,0), e_2 = (0,1,0)$ *e* $e_3 = (0,0,1)$.

1. *Considere a matriz* $A = \begin{bmatrix} 1 & 4 & 7 \\ 1 & 2 & 1 \end{bmatrix}$. *A transformação linear* $A : \mathbb{R}^3 \to \mathbb{R}^2$ *leva os vetores* $\{e_1, e_2, e_3\}$ *da base canônica de* $\mathbb{R}^3$ *nos vetores* $A(e_1) = (1,1)$, $A(e_2) = (4,2)$

e $A(e_3) = (7, 1)$*. Note que os vetores* $(1, 1)$ *e* $(4, 2)$ *não são colineares, logo são LI, e portanto formam uma base de* $\mathbb{R}^2$*. Em particular, o vetor* $A(e_3) = (7, 1)$ *é combinação linear dos vetores* $A(e_1) = (1, 1)$ *e* $A(e_2) = (4, 2)$*.*

2. *Considere a matriz* $A = \begin{bmatrix} 1 & 4 & 6 \\ 1 & 2 & 4 \\ 1 & 1 & 3 \end{bmatrix}$*. Temos então uma transformação linear* $A : \mathbb{R}^3 \to \mathbb{R}^3$*. Observe que* $A(e_1) = (1, 1, 1)$*,* $A(e_2) = (4, 2, 1)$ *e* $A(e_3) = (6, 4, 3)$*. Portanto,*

$$A(e_3) = 2A(e_1) + A(e_2).$$

Isso implica que

$$A(-6, -4, -2) = A(e_3 - 2e_1 - e_2) = A(e_3) - 2A(e_1) - A(e_2) = 0.$$

Em particular, para todo $\lambda \in \mathbb{R}$*, tem-se* $A(-6\lambda, -4\lambda, -2\lambda) = 0$*. Desse modo, a matriz A transforma o plano gerado pelos vetores* e_1 *e* e_2 *no plano gerado pelos vetores* $(1, 1, 1)$ *e* $(4, 2, 1)$ *(note que estes vetores não são colineares), e, ao mesmo tempo, colapsa a reta gerada pelo vetor* $(-6, -4 - 2)$ *no vetor nulo.*

9.2 A matriz de uma transformação linear

Para terminar esse capítulo vamos explicar uma propriedade fundamental das transformações lineares: o fato de que, pondo-se bases no espaço de partida e no espaço de chegada da transformação, podemos associar a ela uma única matriz, de forma que **essencialmente**[2] as transformações lineares que vimos anteriormente são as únicas que existem.

9.2.1 Combinações lineares

Sejam E e F espaços vetoriais de dimensão finita e $f : E \to F$ uma transformação linear. Corroborando a ideia de que uma transformação linear é uma função que respeita a estrutura

[2]Isso porque a matriz da transformação depende apenas das bases nos espaço de partida e no espaço de chegada. Se você muda as bases, a matriz muda.

algébrica dos espaços vetoriais, vamos demonstrar a seguir que toda transformação linear respeita combinações lineares.

Proposição 9.2.1. *Se $x = \sum_{i=1}^{n} \alpha_i y_i$ em E então $f(x) = \sum_{i=1}^{n} \alpha_i f(y)$ em F. Em palavras: se $x \in E$ se escreve como combinação linear dos vetores $y_1, ..., y_n$ de E então sua imagem por f se escreve como combinação linear dos vetores $f(y_1), ..., f(y_n)$ de F, com os mesmo coeficientes.*

Demonstração. A prova se faz por indução no número de vetores envolvidos na combinação linear. Se $n = 1$, o enunciado se reduz a linearidade de f: $f(\alpha_1 x_1) = \alpha_1 f(x_1)$. Agora, suponhamos que o enunciado seja verdadeiro para $n = k$ vetores e vamos provar que ele é verdade para $n = k+1$ vetores. Então, temos $k+1$ vetores $y_1, ..., y_{k+1}$ em E, e $k+1$ números reais $\alpha_1, ..., \alpha_{k+1}$. Seja

$$x = \sum_{i=1}^{k+1} \alpha_i y_i \ \in \ E.$$

Então, escrevendo $x = \sum_{i=1}^{k} (\alpha_i y_i) + \alpha_{k+1} y_{k+1}$ e usando a linearidade de f obtemos[3]

$$f(x) = f(\sum_{i=1}^{k} (\alpha_i y_i) + \alpha_{k+1} y_{k+1}) = f(\sum_{i=1}^{k} \alpha_i y_i) + f(\alpha_{k+1} y_{k+1}). \tag{9.9}$$

Pela linearidade temos que $f(\alpha_{k+1} y_{k+1}) = \alpha_{k+1} f(y_{k+1})$. Pela hipótese de indução, $f(\sum_{i=1}^{k} (\alpha_i y_i) = \sum_{i=1}^{k} \alpha_i f(y_i)$. Usando essas duas igualdades em (9.9) obtemos

$$f(x) = \sum_{i=1}^{k} \alpha_i f(y_i) + \alpha_{k+1} f(y_{k+1}) = \sum_{i=1}^{k+1} \alpha_i f(y_i).$$

Portanto, a propriedade anunciada é verdadeira para $n = k + 1$. Pelo princípio da indução a demonstração está terminada. □

9.2.2 A matriz de f relativa a bases de E e de F

Considere agora $\mathcal{B} = \{b_1, ..., b_d\}$ uma base de E e $\mathcal{B}' = \{v_1, ..., v_n\}$ uma base de F. Para cada $j \in \{1, ..., d\}$ o vetor $f(b_j) \in F$ pode ser escrito como combinação linear dos vetores da base $\mathcal{B}'$; ou seja, existem números reais $a_{1j}, ..., a_{nj}$ tais que

$$f(b_j) = \sum_{i=1}^{n} a_{ij} v_i.$$

[3]Ou seja, usamos $f(a + b) = f(a) + f(b)$ com $a = \sum_{i=1}^{k} (\alpha_i y_i)$ e $b = \alpha_{k+1} y_{k+1}$.

Com essas informações, podemos montar uma matriz $A = [a_{ij}]_{n\times d}$, chamada *matriz de f relativa às bases $\mathcal{B}$ e $\mathcal{B}'$*, cujas colunas são os vetores $c_j = (a_{1j}, ..., a_{nj}) \in \mathbb{R}^n$, e as linhas são os vetores $\ell_i = (a_{i1}, ..., a_{id}) \in \mathbb{R}^d$. As colunas de A possuem ainda uma interpretação interessante:

A j-ésima coluna da matriz de f relativa às bases $\mathcal{B}$ e $\mathcal{B}'$ é o vetor $c_j \in \mathbb{R}^n$ cuja i-ésima entrada é i-ésima coordenada (na base $\mathcal{B}'$) do vetor $f(b_j)$ (o j-ésimo vetor da base $\mathcal{B}$).

Agora, usando a base $\mathcal{B}$ podemos identificar cada vetor $x \in E$ com um vetor $x \in \mathbb{R}^d$: basta escrever $x = \sum_{i=1}^d x_i b_i \in E$ e identificar x com o vetor $(x_1, ..., x_d) \in \mathbb{R}^d$. Similarmente, cada vetor $y \in F$ pode ser identificado com um vetor de $\mathbb{R}^n$ usando a base $\mathcal{B}'$. Com essas identificações, $f : E \to F$ pode ser reinterpretada como uma função $f : \mathbb{R}^d \to \mathbb{R}^n$, basta escrever $f(x) \in \mathbb{R}^n$ como as coordenadas da imagem de x por f em F. Ou seja, se $x = \sum_{i=1}^d x_i b_i$ e $f(x) = \sum_{i=1}^n y_i v_i$ então definimos a função $f : \mathbb{R}^d \to \mathbb{R}^n$ impondo

$$f(x_1, ..., x_d) = (y_1, ..., y_n) \in \mathbb{R}^n.$$

E qual a relação de $f : \mathbb{R}^d \to \mathbb{R}^n$ com a matriz de f?– Vamos provar agora que a $f : \mathbb{R}^d \to \mathbb{R}^n$ é a transformação linear definida pela matriz de f. Com efeito, usando transformação lineares levam combinações lineares em combinações lineares, podemos escrever

$$\begin{aligned} f(x) &= f(\sum_{j=1}^d x_j b_j) = \sum_{j=1}^d x_j f(b_j) \\ &= \sum_{j=1}^d x_j \sum_{i=1}^n a_{ij} v_i \\ &= \sum_{j=1}^d \sum_{i=1}^n x_j a_{ij} v_i \\ &= \sum_{i=1}^n \left(\sum_{j=1}^d x_j \right) v_i \\ &= \sum_{i=1}^n \langle x, \ell_i \rangle v_i \end{aligned}$$

Mais exemplos de transformações lineares

Função antípoda– Considere $f : \mathbb{R}^d \to \mathbb{R}^d$ dada por $f(x) = -x$. Como não estamos fazendo nenhuma menção às bases usadas, vamos sempre supor tacitamente que estamos considerando a base canônica $\{e_1, ..., e_d\}$ de $\mathbb{R}^d$. Como $f(e_j) = -e_j$ concluímos que a j-ésima coluna da matriz de f é o vetor $-e_j$. Portanto, a matriz de f é $-\mathrm{I}$.

Inversão de coordenadas– Considere $f : \mathbb{R}^d \to \mathbb{R}^d$ dada por $f(x_1, ..., x_d) = (x_d, ..., x_1)$. Então $f(e_j) = e_{d-j}$, e portanto a matriz de f é

$$A = \begin{bmatrix} 0 & \dots & 1 \\ \vdots & \iddots & \vdots \\ 1 & \dots & 0 \end{bmatrix}$$

Soma corrente– Considere $f : \mathbb{R}^d \to \mathbb{R}^d$ dada por

$$f(x_1, ..., x_d) = (x_1, x_1 + x_2, x_1 + x_2 + x_3, ..., \sum_{\ell=1}^{d} x_\ell)$$

Note que $f(e_1) = \mathbb{I}$, $f(e_2) = (0, 1, 1, ..., 1)$, $f(e_3) = (0, 0, 1, 1, ..., 1)$,..., $f(e_d) = e_1$. Assim, a matriz de f na base canônica de $\mathbb{R}^d$ é

$$A = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

Operador vetor centrado– Considere $f : \mathbb{R}^d \to \mathbb{R}^d$ dada por $f(x) = x - \mu(x)\mathbb{I}$. A matriz de f é

$$A = \begin{bmatrix} 1 - \frac{1}{d} & -\frac{1}{d} & \dots & -\frac{1}{d} \\ -\frac{1}{d} & 1 - \frac{1}{d} & \dots & -\frac{1}{d} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{d} & -\frac{1}{d} & \dots & 1 - \frac{1}{d} \end{bmatrix}$$

9.2.3 Exemplos de aplicações

Em geral as funções que aparecem em aplicações e em modelos do mundo real não são lineares (nem sequer afins). No entanto, quase sempre elas podem ser *aproximadas* em algum sentido por funções lineares. Dessa forma modelos mais realísticos (e por isso mesmo mais complexos) podem ser estudos de forma aproximada por modelos muito mais simples.

Elasticidade preço-demanda

Considere um conjunto de n bens ou serviços cujos preços são dados por em vetor $p \in \mathbb{R}^n$ e suponha que a demanda por cada um desses produtos seja dada por um vetor $d \in \mathbb{R}^n$. Uma variação nos preços induz uma variação na demanda e gostaríamos de entender de que forma a variação na demanda depende da variação nos preços.

Vamos considerar que o novo vetor de preços da economia seja $\overline{p} \in \mathbb{R}^n$. O vetor que dá a variação percentual de preços é o vetor $\delta^p \in \mathbb{R}^n$ cuja j-ésima coordenada é

$$\delta^p_j \stackrel{\text{def}}{=} \frac{\overline{p}_j - p_j}{p_j}$$

Similarmente, denotando por $\overline{d} \in \mathbb{R}^n$ o novo vetor de demanda, o vetor que fornece a variação percentual de demanda é o vetor $\delta^d \in \mathbb{R}^n$ cuja j-ésima entrada é

$$\delta^d_j \stackrel{\text{def}}{=} \frac{\overline{d}_j - d_j}{d_j}$$

Com essas notações, o problema que queremos resolver é: de que forma o vetor δ^d depende do vetor δ^p? Um modelo que responda a essa pergunta é chamada *modelo de elasticidade preço demanda*. Num modelo de elasticidade linear, a variação na demanda $\delta^d \in \mathbb{R}^n$ se relaciona com a variação nos preços $\delta^p \in \mathbb{R}^n$ através de uma matriz $E = [\theta_{ij}]_{n\times n}$:

$$\delta^d = E(\delta^p).$$

Por exemplo, suponha que $\theta_{11} = -0.4$. Então, se aumentamos em 1% o preço do primeiro bem, deixando os outros fixos, vamos ver um decréscimo de 0.4% na demanda desse mesmo bem. Por outro lado, se $\theta_{21} = 0.2$ então, de acordo com esse modelo, um aumento de 1% no preço do primeiro bem vai causar um aumento de 0.2% na demanda pelo segundo bem. Nesse caso, os economistas dizem que o segundo bem funciona como um *substituto parcial* do primeiro.

Deformações estruturais

Considere uma estrutura de aço como uma ponte ou a estrutura interna de um prédio. Seja f um vetor de $\mathbb{R}^n$ que dá a força exercida na estrutura em n pontos diferentes dela. A estrutura vai se deformar levemente devido a essas forças. Seja $d \in \mathbb{R}^n$ o vetor que dá o deslocamento da estrutura em cada ponto. Para pequenos deslocamentos a relação entre eles e a força sofrida pela estrutura é bem aproximada por um modelo linear:

$$d = A(f),$$

onde $A = [a_{ij}]_{n\times n}$ é uma matriz.

9.3 Funções afins

Seja $A = [a_{ij}]_{d\times d}$ uma matriz. Vamos denotar por $A : \mathbb{R}^d \to \mathbb{R}^d$ a função linear cuja matriz na base canônica é A. Uma *função afim* é uma função $f : \mathbb{R}^d \to \mathbb{R}^d$ da forma

$$f(x) = A(x) + b,$$

onde $b \in \mathbb{R}^d$ é um vetor fixado. As funções afins, ao contrário das lineares, não preservam combinações lineares. Contudo, elas preservam combinações afins. Com efeito, se α e β são números reais que somam 1 e se x, y são vetores de $\mathbb{R}^d$ então, como $\alpha + \beta = 1$ podemos escrever $b = (\alpha + \beta)b = \alpha b + \beta b$. Logo,

$$\begin{aligned} f(\alpha x + \beta y) &= A(\alpha x + \beta y) + b \\ &= \alpha A(x) + \beta A(y) + \alpha b + \beta b \\ &= \alpha\left(A(x) + b\right) + \beta\left(A(y) + b\right) \\ &= \alpha f(x) + \beta f(y) \end{aligned}$$

Capítulo 10

Exemplos e aplicações de matrizes

Informados da íntima relação entre matrizes e transformações lineares entre espaços vetoriais de dimensão finita, vamos agora nos deter no estudo mais sistemático de matrizes. Com o que acabamos de ver, sabemos que isso não nos tira do objetivo guia, que é enxergara estrutura algébrica dos espaços vetoriais através das propriedades geométricas das transformações lineares entre eles.

10.1 Aplicações

Imagens

Uma imagem preto e branca com $m \times n$ pixeis pode ser interpretada como uma matriz. Nesse caso, a entrada a_{ij} representa a intensidade de cor do pixel na posição vertical i e na posição horizontal j.

Dados climáticos

Uma matriz $m \times n$ pode ser usada para sintetiza a informação pluviométrica ao longo de n dias consecutivos em m localidades diferentes. Nesse contexto, por exemplo, a entrada a_{34} fornece a quantidade de chuva na 3º localidade ao longo 4º dia.

Tabela de contingência

Suponha uma coleção de objetos que podem possuir dois atributos, sendo que existem m possibilidades para o primeiro atributo (numeradas de 1 a m) e n possibilidades para o segundo (numeradas de 1 a n). Podemos formar uma matriz $A = [a_{ij}]_{m\times n}$ na qual a entrada a_{ij} fornece a quantidade de objetos cujo primeiro atributo é igual a i e cujo segundo é igual a j. Por exemplo, para cada estudante da educação básica no Brasil podemos atribuir duas características: a série escolar em que ele se encontra (numeradas de 1 a 12) e o estado da federação onde ele mora (numerados de 1 a 27). Nesse caso, a entrada $a_{11,17}$ fornece a quantidade estudantes no 2º ano do ensino médio (penúltimo ano de formação na educação básica) no estado 17 da federação.

10.2 Tipos especiais de matrizes

Matriz de zeros

É a matriz $Z_{m\times n} = [0]_{m\times n}$ que possui todas as entradas nulas. Claramente a transformação linear $Z : \mathbb{R}^n \to \mathbb{R}^m$ satisfaz $Z(x) = 0$ para todo $x \in \mathbb{R}^n$.

Matriz identidade

É a matriz $\mathrm{I} = [\delta_{ij}]$, onde $\delta_{ij} = 1$ se $i = j$ e 0 caso contrário. Ou seja

$$\mathrm{I} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}.$$

Matriz esparsa

Uma matriz é esparsa quando muitas de suas entradas são nulas. A fração das entradas que não é nula é chamada o padrão de esparsidade da matriz.

Matriz diagonal

Uma matriz diagonal é qualquer matriz cujas entradas a_{ij} satisfazem $i \neq j \implies a_{ij} = 0$. Obviamente uma matriz diagonal é esparsa.

Matriz triangular

$A = [a_{ij}]$ é triangular superior se $i > j \implies a_{ij} = 0$.

A transposta

A transposta de uma matriz A é matriz cujas linhas são as colunas de A. É denotada por A^T.

Matrizes simétricas

Uma matriz é simétrica se $A = A^T$

10.3 Algumas transformações geométricas

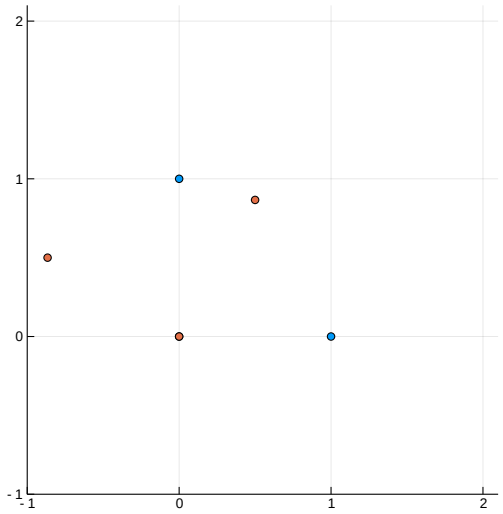

Figura 10.1: Rotação de 60º de um triângulo retângulo.

10.4 Álgebra de matrizes

Vamos denotar por $\mathcal{M}(m \times n)$ o conjunto de todas as matrizes $m \times n$. Podemos somar dois elementos $A, B \in \mathcal{M}(m \times n)$, pondo

$$A + B \stackrel{\text{def}}{=} [a_{ij} + b_{ij}]_{m \times n}.$$

Similarmente, podemos multiplicar $A \in \mathcal{M}(m \times n)$ por um número real $\lambda \in \mathbb{R}$ pondo

$$\lambda A \stackrel{\text{def}}{=} [\lambda a_{ij}]_{m \times n}.$$

Com essas operações o conjunto $\mathcal{M}(m \times n)$ fica dotado de uma estritura de espaço vetorial. Uma outra forma de ver isso é interpretar uma matriz como um vetor de $\mathbb{R}^{mn}$, ou seja, pensar numa matriz como uma lista ordenada de mn números. Em particular, isso deixa claro que os espaço das matrizes tem dimensão finita.

10.4.1 Medindo o comprimento de uma matriz

Sendo $\mathcal{M}(m \times n)$ um espaço vetorial, podemos dotá-lo de diversas normas diferentes.

Norma euclideana

$\|A\|_2 \stackrel{\text{def}}{=} \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2}$

Norma de operador

$\|A\| = \sup\{\|A(x)\|; \|x\| = 1\}$

Capítulo 11

Grafos e matrizes I

Um grafo é um objeto matemático simples, composto por vértices e arestas que conectam os vértices.

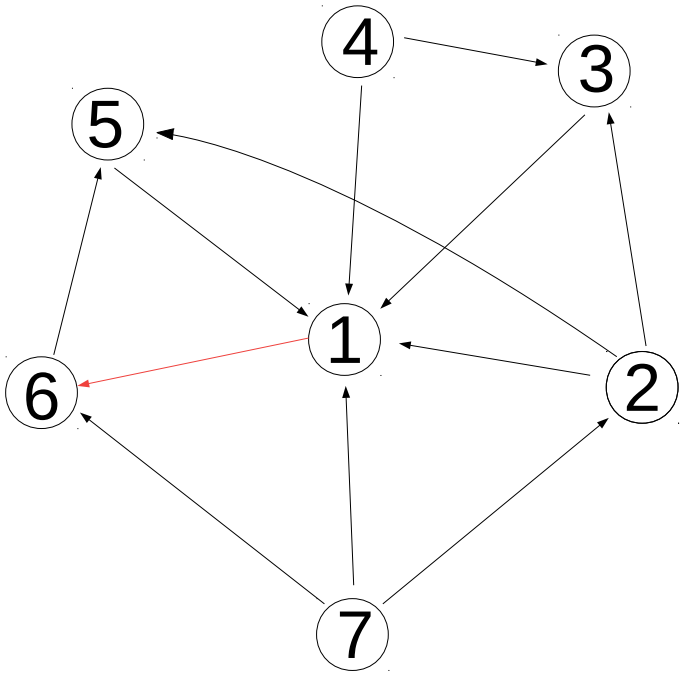

Figura 11.1: A estrutura de *links* da internet pode ser descrita através de um grafo.

Grafos podem ser usados para modelar todo tipo de relacionamento e as aplicações da

Teoria dos Grafos às vezes são de cair o queixo. Um exemplo delas é o *Google*. A empresa começou lá na década de 90 com um algoritmo que propôs uma forma mais matematizada de se fazer buscas na internet. E a base desse algoritmo é justamente um teorema de álgebra linear. Escrevi um artigo sobre isso aqui `http://www.professores.uff.br/brunosantiago/vulgarisation-divulgacao-da-matematica/`.

Neste capítulo vamos ver como podemos associar matrizes a um grafo e como elas nos permitem descobrir propriedades maravilhosas sobre os grafos.

Vamos começar falando sobre grafos de maneira formal.

Definição 11.0.1 (Grafos)**.** *Um grafo é composto por um conjunto finito $\mathcal{V} = \{v_1, ..., v_n\}$, chamado conjunto de vértices do grafo, e um subconjunto $\mathcal{A} \subset \mathcal{V} \times \mathcal{V}$ do produto cartesiano, chamado conjunto de arestas do grafo.*

Por exemplo, se $\mathcal{V} = \{1, 2, 3\}$ possui três elementos, e o conjunto de arestas for $\mathcal{A} = \{(1, 2), (2, 1), (1, 3), (2, 3)\}$ teremos o grafo desenhado abaixo.

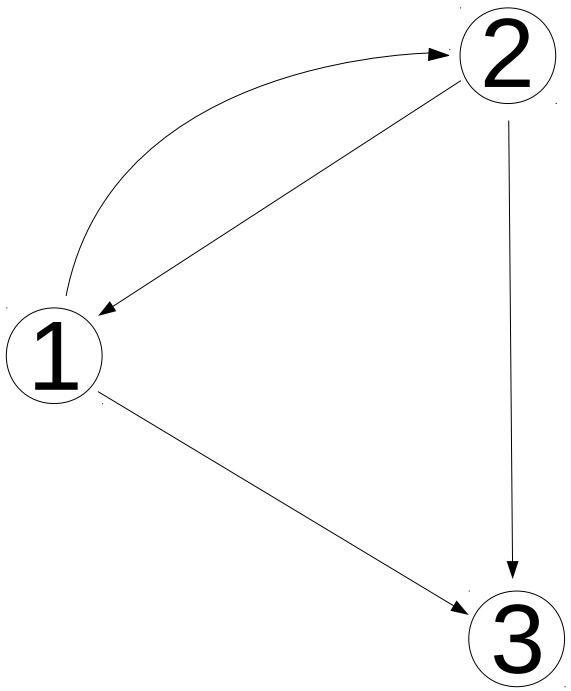

Figura 11.2: Exemplo de um grafo.

11.1 A matriz de adjacência de um grafo

Considere $\mathcal{G}$ um grafo com n vértices. Podemos encapsular a informação dada pelo conjunto de arestas $\mathcal{A}$ de $\mathcal{G}$ através da *matriz de adjacência* de $\mathcal{G}$.

Definição 11.1.1 (Matriz de Adjacência)**.** *Seja $\mathcal{G}$ um grafo com conjunto de vértices $\mathcal{V} = \{1, ..., n\}$, contendo n elementos, e com conjunto de arestas $\mathcal{A}$. Para cada $i, j \in \{1, ..., n\}$ seja*

$$a_{ij} = \begin{cases} 1, \text{ se } (i,j) \in \mathcal{A} \\ 0, \text{ caso contrário} \end{cases}.$$

A matriz $A = [a_{ij}]_{n \times n}$ é chamada matriz de adjacência do grafo A.

Exemplo 11.1.2. *A matriz de adjacência do grafo da Figura 11.2 é*

$$A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Exemplo 11.1.3. *A matriz de adjacência do grafo da Figura 11.1 é*

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

11.2 A matriz de incidência de um grafo

Nesta seção vamos considerar apenas *grafos dirigidos:* um grafo é dirigido se

$$(i,j) \in \mathcal{A} \implies (j,i) \notin \mathcal{A}.$$

Definição 11.2.1 (Matriz de incidência)**.** *Seja $\mathcal{G}$ um grafo com conjunto de vértices $\mathcal{V} = \{1, ..., n\}$ e conjunto de arestas $\mathcal{A}$. Para cada $i, j \in \{1, ..., n\}$ considere*

$$g_{ij} = \begin{cases} 1, \ \ se \ (j,i) \in \mathcal{A} \\ -1, \ \ se \ (i,j) \in \mathcal{A} \\ 0, \ \ caso \ contrário \end{cases}$$

A matriz $G = [g_{ij}]_{n \times n}$ é chamada matriz de incidência *do grafo $\mathcal{G}$.*

11.3 A energia de Dirichlet de um grafo

Capítulo 12

Núcleo e Imagem

Nesta seção, iremos olhar os exemplos anteriores de um ponto de vista mais conceitual. Já vimos que uma transformação linear (matrizes em particular) sempre transforma subespaços em subespaços. Vimos também diversos casos onde retas são transformadas em retas, planos em planos e até mesmo planos e retas sendo colapsados no zero. Para tratar esses casos particulares de um ponto de vista mais geral, faremos as seguintes definições:

12.1 Definição

Definição 12.1.1. *Seja $A : E \to F$ uma transformação linear entre os espaços vetoriais E e F. O* núcleo *de A, denotado por $N(A)$, é o conjunto dos vetores em E que são transformados por A no zero de F. Em símbolos:*

$$N(A) = \{v \in E; A(v) = 0\}.$$

A imagem *de A, denotada por $Im(A)$, é conjunto dos vetores de F que são "atingidos"pela transformação:*

$$Im(A) = \{u \in F; \exists\, v \in E; u = A(v)\}.$$

Teorema 12.1.2. *Seja $A : E \to X$ uma transformação linear entre os espaços vetoriais E e X. Se $F \subset E$ é um subespaço, então $A(F) = \{A(v) \in X; v \in F\}$ é um subespaço de X.*

Demonstração. Como $0 \in F$, e $0 = A(0)$, segue que $0 \in A(F)$. Além disso, se $A(v), A(u) \in F$ então $A(u) + A(v) = A(u + v) \in A(F)$. Analogamente, se $A(v) \in A(F)$ e $\lambda \in \mathbb{R}$, temos que $\lambda A(v) = A(\lambda v) \in A(F)$. □

Proposição 12.1.3. *$N(A)$ é um subespaço vetorial de E; $Im(A)$ é um subespaço vetorial de F.*

Demonstração. Sejam $u, v \in N(A)$. Então

$$A(u + v) = A(u) + A(v) = 0 + 0 = 0,$$

logo $u + v \in N(A)$. Se $v \in N(A)$ e $\lambda \in \mathbb{R}$, então

$$A(\lambda v) = \lambda A(v) = \lambda 0 = 0,$$

e portanto $\lambda v \in N(A)$.

Sejam $A(u), A(v) \in Im(A)$. Então,

$$A(u) + A(v) = A(u + v) \in Im(A).$$

Analogamente, se $\lambda \in \mathbb{R}$ e $A(v) \in Im(A)$ então

$$\lambda A(v) = A(\lambda v) \in Im(A).$$

□

O núcleo de uma transformação linear nos dá informação sobre quantas dimensões ela colapsa. Em particular, quando o núcleo tem o menor tamanho possível, a transformação não deleta informação.

Proposição 12.1.4. *Seja $f : X \to Y$ uma transformação linear entre espaços vetoriais X e Y. Suponha que $\dim X = d$ e seja $\{x_1, ..., x_d\}$ uma base de X. Se $N(f) = \{0\}$ então $\{f(x_1), ..., f(x_d)\}$ é uma base de $Im(f)$.*

Demonstração. Precisamos provar que $\{f(x_1), ..., f(x_d)\}$ é um gerador LI de Y. Vamos começar vendo que é LI. Considere uma combinação linear nula qualquer

$$\sum_{l=1}^{d} \alpha_l f(x_l) = 0.$$

Como f é linear, para cada l temos $\alpha_l f(x_l) = f(\alpha_l x_l)$ e $\sum_{l=1}^{d} f(\alpha_l x_l) = f(\sum_{l=1}^{d} \alpha_l x_l)$. Concluímos que

$$f\left(\sum_{l=1}^{d} \alpha_l x_l\right) = 0,$$

o que mostra que $\sum_{l=1}^{d} \alpha_l x_l \in N(f)$. Como assumimos que $N(f) = \{0\}$, somos levados a concluir que $\sum_{l=1}^{d} \alpha_l x_l = 0$. No entanto, $\{x_1, ..., x_d\}$ é um conjunto LI em X e dessa forma temos que $\alpha_l = 0$, $\forall\, l = 1, ..., d$. Isso mostra que $\{f(x_1), ..., f(x_d)\}$ é LI.

Tome $y \in Im(f)$ arbitrário. Então, existe $x \in X$ tal que $y = f(x)$. Como $\{x_1, ..., x_d\}$ é uma base de X, existem números $\alpha_1, ..., \alpha_d$ tais que

$$x = \sum_{l=1}^{d} \alpha_l x_l.$$

Como f é linear, isso nos leva a

$$f(x) = f\left(\sum_{l=1}^{d} \alpha_l x_l\right) = \sum_{l=1}^{d} \alpha_l f(x_l),$$

e portanto y é combinação linear dos vetores $\{f(x_1), ..., f(x_d)\}$. Isso mostra que o conjunto é também gerador de $Im(f)$ conclui a prova. □

O núcleo e a imagem de uma transformação linear são relacionados pelo seguinte teorema fundamental:

Teorema 12.1.5. *Seja $f : X \to Y$ uma função linear entre espaços vetoriais X e Y. Então,*

$$\dim N(f) + \dim Im(f) = \dim X.$$

Demonstração. Como $N(f)$ é um subespaço e portanto é espaço vetorial em si próprio, podemos tomar uma base $\mathcal{B} = \{b_1, ..., b_k\}$ de $N(f)$. Em particular $\dim N(f) = k$. Além disso, por ser uma base de um subespaço co conjunto $\mathcal{B}$ é LI em X. Portanto podemos completar esse conjunto de modo a obter uma base. Ou seja, existem vetores $v_1, ..., v_\ell$ de forma que

$$\{b_1, ..., b_k, v_1, ..., v_\ell\}$$

é uma base de X. Em particular, isso significa que $\dim X = k + \ell$. Vamos provar que $\{f(v_1), ..., f(v_\ell)\}$ é uma base para o subespaço $Im(f)$ de Y. Com efeito, se $x \in X$ então existem números $\alpha_1, ..., \alpha_k$ e $\beta_1, ..., \beta_\ell$ tais que

$$x = \alpha_1 b_1 + ...\alpha_k b_k + \beta_1 v_1 + ... + \beta_\ell v_\ell = b + v,$$

onde $b = \sum_{i=1}^{k} \alpha_i b_i \in N(f)$ e $v = \sum_{i=1}^{\ell} \beta_i v_i$. Pela linearidade, como $x = b + v$, $f(x) = f(b) + f(v)$. Como $b \in N(f)$, $f(b) = 0$. Isso prova que $f(x) = f(v)$. Aplicando a linearidade mais uma vez concluímos,

$$f(v) = \sum_{i=1}^{\ell} \beta_i f(v_i).$$

Isso prova que o conjunto $\{f(v_1), ..., f(v_\ell)\}$ é gerador de $Im(f)$. Suponha agora que $\beta_1, ..., \beta_\ell$ sejam números reais tais que

$$\sum_{i=1} \beta_i f(v_i) = 0.$$

Pela linearidade isso implica que $f(v) = 0$, logo $v \in N(f)$. Mas o único vetor gerado por $\{v_1, ..., v_\ell\}$ é o vetor nulo. Logo $v = 0$. Como $\{v_1, ..., v_\ell\}$ é LI, concluímos que $\beta_i = 0$ para todo i. Isso prova que o conjunto $\{f(v_1), ..., f(v_\ell)\}$ é LI. Logo, $\{f(v_1), ..., f(v_\ell)\}$ é uma base de $Im(f)$ e portanto $\dim Im(f) = \ell$. Como $\dim X = k + \ell$, e $k = \dim N(f)$, isso prova exatamente a fórmula anunciada. □

A dimensão do núcleo mais a dimensão da imagem é igual a dimensão do espaço de partida da transformação. Vai levar um certo tempo, caro leitor, até que possamos apreciar toda a profundidade da fórmula e de sua bela demonstração que acabamos de ver. Vamos discutir a seguir uma aplicação não-trivial disso tudo que pode ajudar nesse processo.

Capítulo 13

Funcionais lineares e aplicações

Vamos agora aplicar o material visto até aqui em economia.

13.1 Programação Linear

Considere uma função linear $f : \mathbb{R}^2 \to \mathbb{R}$. Então, como a imagem é um subespaço de $\mathbb{R}$, ou $\dim Im(f) = 0$, e nesse caso a função é completamente trivial (anula todos os pontos de $\mathbb{R}^2$) ou então $\dim Im(f) = 1$ e, nesse caso, pelo Teorema do Núcleo e da Imagem, $\dim N(f) = 1$. Vamos nos concentrar nesse caso. Apenas para efeitos didáticos e ilustrativos, vamos pensar a função f como uma distribuição de temperaturas no plano. Assim, cada ponto do plano corresponde a uma temperatura. Em particular, os pontos do núcleo de f correspondem à temperatura zero (veja a Figura 13.1 a seguir). Sendo um subespaço unidimensional do plano, $N(f)$ é uma reta que passa pela origem. Na nossa analogia, todos os pontos dessa reta estão a temperatura zero. Sabemos que f dá zero somente no núcleo. Então, se $v \in \mathbb{R}^2 \backslash N(f)$ devemos ter $f(v) \neq 0$. Mais do que isso, como $\dim Im(f) = 1$, sabemos que $Im(f) = \mathbb{R}$ (ou seja, f é sobrejetiva). Então, para todo número real existe um vetor do plano cuja imagem é esse número. Ou, todas as temperaturas podem ser sentidas em algum lugar do plano. Então, tomamos um vetor v tal que $f(v) = 1$. Considere agora uma temperatura $t \in \mathbb{R}$ e vamos olhar a reta $r = N(f) + tv$. Ou seja,

$$r = \{p \in \mathbb{R}^2; \exists\, u \in N(f); p = u + tv\}.$$

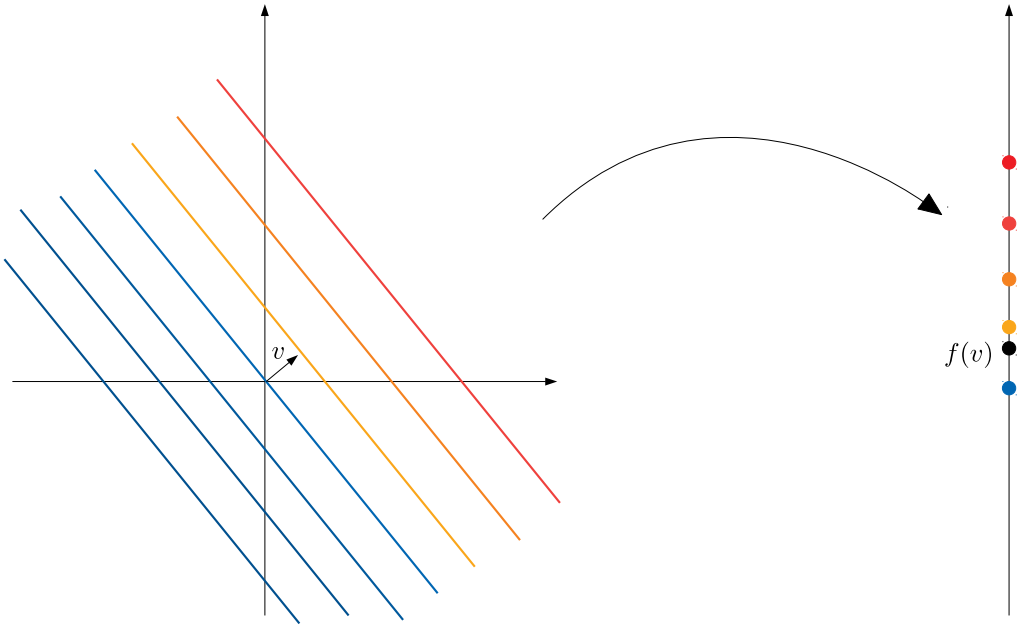

Figura 13.1: A reta azul é o núcleo de f. Cada reta paralela é levada em único ponto (correspondem a regiões de temperatura constante).

Seja $p \in r$ (você pode visualizar, por exemplo, r como sendo a reta vermelha na Figura 13.1). Então, existe $u \in N(f)$ tal que $p = u + tv$ e portanto

$$f(p) = f(u + tv) = f(u) + tf(v) = 0 + t.1 = t.$$

Ou seja, em toda a reta r todos os pontos têm a mesma temperatura! No jargão puramente matemático, diríamos "f é constante igual a t ao longo da reta r". Mas, seriam os pontos de r os únicos com essa temperatura? De fato, se $p \in \mathbb{R}^2$ é tal que $f(p) = t$ então, como $f(v) = 1$ podemos escrever $f(p) = tf(v)$ e portanto $f(p - tv) = 0$, donde concluímos que existe $u \in N(f)$ tal que

$$u = p - tv \iff p = u + tv \in r.$$

Logo, a resposta é sim: a reta $r = N(f) + tv$ é único lugar do plano onde se tem temperatura igual a t. Ou, na linguagem matemática técnica

$$r = f^{-1}(t) \stackrel{\text{def}}{=} \{p \in \mathbb{R}^2; f(p) = t\}.$$

Isso nos mostra que a temperatura é constante nas retas paralelas ao núcleo de f, e ela aumenta quando caminhamos paralelamente ao vetor v e ela diminui se caminhamos paralelamente a $-v$. Por isso, na Figura 13.1 as retas paralelas ao núcleo que estão no sentido

positivo de v vão ficando cada mais vermelhas (a temperatura aumenta), enquanto aquelas que estão no sentido negativo de v vão ficando cada vez mais azuis (a temperatura diminui).

A beleza dessa discussão é que ela se aplica a **qualquer** função linear não-constante $f : \mathbb{R}^2 \to \mathbb{R}$. Sempre podemos identificar as retas onde a função é constante e o sentido de crescimento/decrescimento da função. Além disso, observe que f colapsa cada reta paralela em único ponto de $\mathbb{R}$ e que se identificamos os pontos em cada reta podemos enxergar f como uma função de $\mathbb{R}$ em $\mathbb{R}$, exatamente como fizemos na demonstração do Teorema do Núcleo e da Imagem. Podemos tirar proveito dessa generalidade para resolver os mais diversos

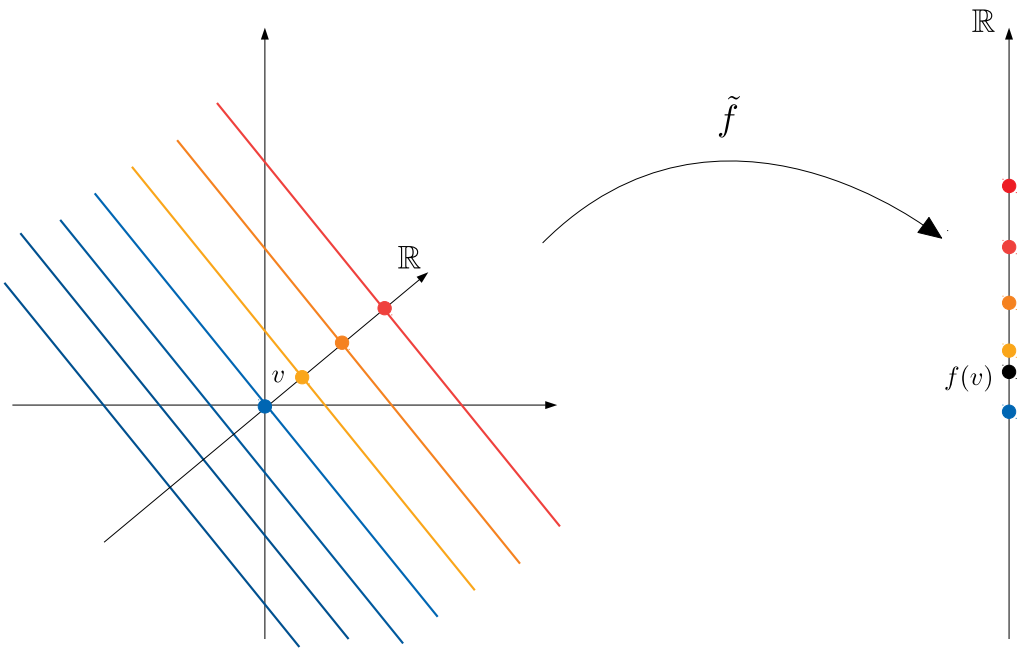

Figura 13.2: Analogia entre a prova do Teorema do Núcleo e da Imagem e a discussão acima: as classes de equivalência são as retas $N(f) + tv$ e f passa ao quociente como uma bijeção $\tilde{f} : \mathbb{R} \to \mathbb{R}$.

problemas. Veremos alguns deles a seguir:

Problema 2. *Um fabricante de doces possui 130Kg de trufas de chocolate com calda de cereja e 170Kg de trufas de chocolate com menta em estoque. Para a páscoa, ele decide vender esse estoque em dois kits promocionais, ambos vendidos a kilo. O primeiro é composto por metade de trufas de cereja e metade de trufas de menta e é vendido a R$ 2,00 o kilo. O segundo é composto por um terço de trufas de cereja e dois terços de trufas de menta e será vendido*

por R$ 1,75 o kilo. Quantos quilos de cada kit devem ser vendidos de modo que a receita seja a maior possível?

Solução. Seja x a quantidade de kilos vendidos no kit 01 e y a quantidade de quilos vendidos no kit 02. A receita oriunda dessa venda é dada pela função

$$f(x, y) = 2x + 1,75y.$$

Observe que essa expressão define uma função linear $f : \mathbb{R}^2 \to \mathbb{R}$. Nosso objetivo nesse problema é maximizar essa função. No entanto, temos algumas restrições para os valores possíveis de serem assumidos pelas variáveis x e y. Em primeiro lugar, esses valores devem ser positivos (≥ 0). Em segundo lugar, a matéria prima usada em cada kit é finita (130 Kg de cereja e 170Kg de menta). A quantidade de trufas de cereja vendida nos dois kits, em quilos, é $x/2 + y/3$ e a quantidade de trufas de menta é $x/2 + 2y/3$. Portanto, devemos ter

$$\varphi(x, y) \stackrel{\text{def}}{=} \frac{x}{2} + \frac{y}{3} \leq 130 \quad \text{e} \quad \psi(x, y) \stackrel{\text{def}}{=} \frac{x}{2} + \frac{2y}{3} \leq 170.$$

Ou seja, a modelagem matemática da situação nos trouxe o seguinte problema: *encontrar o ponto em* $p = (x, y) \in \mathbb{R}^2$*, sujeito às restrições*

$$\begin{aligned} x &\geq 0 \\ y &\geq 0 \\ \varphi(x, y) &\leq 130 \\ \psi(x, y) &\leq 170, \end{aligned}$$

no qual a função $f : \mathbb{R}^2 \to \mathbb{R}$ *assume o seu maior valor.* Assim como f, as funções φ e ψ são lineares. O núcleo de φ é a reta de equação $y = -\frac{3x}{2}$. Essa reta é gera pelo vetor $(1, -3/2)$. Já o núcleo da função ψ é a reta de equação $y = -\frac{3}{4}$, gerada pelo vetor $(1, -3/4)$. □

Capítulo 14

A transposta

Neste capítulo vamos demonstrar um fato clássico da álgebra linear, a igualdade entre o posto linha e o posto coluna. Este teorema diz que se a gente olha o subespaço gerado pelas linhas de uma matriz, e olha o subespaço gerado pelas colunas de uma matriz, ainda que possam ser subespaços diferentes, eles possuem a **mesma dimensão.**

Esse é um fato extremante útil na hora de fazer contas e na hora de escrever algorítimos mais econômicos. Muitas vezes, por causa desse teorema, é possível resolver dois problemas com um mesmo algorítimo.

Além da importância prática da igualdade entre o posto linha e o posto coluna, neste capítulo vamos aplicar uma estratégia de estudo muito particular da matemática contemporânea, que consiste em olhar como um objeto se modifica quando submetido a determinadas transformações. As mudanças sofridas pela objeto podem revelar padrões intrínsecos, que não seriam postos em evidência de outra forma.

14.1 A matriz transposta

Seja $f : \mathbb{R}^d \to \mathbb{R}^k$ uma função linear, representada na base canônica pela matriz $k \times d$

$$F = \begin{bmatrix} a_{11} & \dots & a_{1d} \\ \vdots & \dots & \vdots \\ a_{k1} & \dots & a_{kd} \end{bmatrix}.$$

Definição 14.1.1. *A* transposta *de f é a função linear $f^T : \mathbb{R}^k \to \mathbb{R}^d$ representada na base canônica pela matriz $F^T = [a_{ji}]_{d\times k}$. Ou seja, as linhas de F^T são as colunas de F.*

Olhar a transposta de f é uma forma de "transformar" f em outra coisa e ver o que acontece, que relações essa outra coisa guarda com f, no intuito de extrair propriedades da função original f. A relação mais fundamental entre f e f^T é dada no teorema a seguir.

Teorema 14.1.2. *Para todo $x \in \mathbb{R}^d$ e para todo $y \in \mathbb{R}^k$, vale que*

$$\langle x, f^T(y)\rangle = \langle f(x), y\rangle.$$

Demonstração. Sejam $\ell_1 = (a_{11}, ..., a_{1d}) \in \mathbb{R}^d$,...,$\ell_k = (a_{k1}, ..., a_{kd}) \in \mathbb{R}^d$ as linhas da matriz F. Então,

$$f(x) = \begin{bmatrix} \langle \ell_1, x\rangle \\ \vdots \\ \langle \ell_k, x\rangle \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^d x_j a_{1j} \\ \vdots \\ \sum_{j=1}^d x_j a_{kj} \end{bmatrix} \in \mathbb{R}^k,$$

e portanto

$$\langle y, f(x)\rangle = \sum_{i=1}^k \sum_{j=1}^d y_i x_j a_{ij}.$$

Sejam agora $c_1 = (a_{11}, ..., a_{k1}) \in \mathbb{R}^k$,...,$c_d = (a_{1d}, ..., a_{kd}) \in \mathbb{R}^k$ as colunas da matriz F, que são as linhas da matriz F^T. Analogamente ao que fizemos para f, temos que

$$f^T(y) = \begin{bmatrix} \langle c_1, y\rangle \\ \vdots \\ \langle c_d, y\rangle \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^k y_i a_{i1} \\ \vdots \\ \sum_{i=1}^k y_i a_{d1} \end{bmatrix} \in \mathbb{R}^d,$$

e portanto

$$\langle x, f^T(y)\rangle = \sum_{i=1}^k \sum_{j=1}^d y_i x_j a_{ij},$$

de onde a igualdade procurada segue imediatamente. □

Definição 14.1.3. *O **posto linha** de f é o posto de sua transposta.*

Como o posto de uma função linear é a dimensão de sua imagem, e como a imagem é o subespaço gerado pelas colunas da matriz, normalmente nos referimos ao posto de uma função linear como o **posto-coluna.** Como as colunas de f^T são as linhas de f, vemos que o posto-linha de f é a dimensão do subespaço gerado pelas suas colunas. Portanto, a igualdade entre o posto-linha e o posto-coluna de uma matriz é a afirmação de que f e f^T possuem o mesmo posto. Na seção a seguir vamos ver como o conceito geométrico de ortogonalidade produz relações surpreendentes entre f e sua transposta f^T, que acabam por demonstrar justamente isso.

14.2 Transposição e ortogonalidade

Sejam $f : \mathbb{R}^d \to \mathbb{R}^k$ uma função linear e $f^T : \mathbb{R}^k \to \mathbb{R}^d$ sua transposta. Lembre que $N(f)$ e $Im(f^T)$ são subespaços de $\mathbb{R}^d$ ao passo que $N(f^T)$ e $Im(f)$ são subespaços de $\mathbb{R}^k$.

Proposição 14.2.1. *(a)* $N(f^T) \subset Im(f)^\perp$

(b) $Im(f^T) \subset N(f)^\perp$

Demonstração. Vamos começar provando o item (a). Devemos provar que todo elemento de $N(f^T)$ é também um elemento de $Im(f)^\perp$. Ou seja, queremos provar que todo vetor no núcleo de f^T é ortogonal a qualquer vetor da imagem de f. Seja portanto $y \in N(f^T)$ e tome $z \in Im(f)$ qualquer. Como $z \in Im(f)$, existe $x \in \mathbb{R}^d$ tal que $z = f(x)$. Assim, aplicando o Teorema 14.1.2 temos que

$$\langle y, z\rangle = \langle x, f(x)\rangle = \langle f^T(y), x\rangle.$$

Como $y \in N(f^T)$, devemos ter $f^T(y) = 0$ e portanto

$$\langle f^T(y), x\rangle = \langle 0, x\rangle = 0,$$

ou seja, $\langle z, y\rangle = 0$. Isso prova que y é ortogonal a todo $z \in Im(f)$, portanto que $y \in Im(f)^\perp$.

A prova do item (b) segue os mesmos passos. Tome $z \in Im(f^T)$ e seja $x \in N(f)$ qualquer. Vamos provar que z é ortogonal a x. Como $z \in Im(f^T)$ existe $y \in \mathbb{R}^k$ tal que $z = f^T(y)$. Portanto,

$$\langle z, x\rangle = \langle f^T(y), x\rangle = \langle y, f(x)\rangle.$$

Como $x \in N(f)$, temos $f(x) = 0$ e como antes isso prova que $\langle z, x\rangle = 0$ e completa o argumento. □

Como consequência da proposição acima, vamos demonstrar a igualdade entre o posto linha e posto coluna. Para simplificar a notação sejam $p = \dim Im(f)$ e $p^T = \dim Im(f^T)$. Vamos denotar também por $n = \dim N(f)$ a nulidade de f e por $n^T = \dim N(f^T)$ a nulidade da transposta de f.

Teorema 14.2.2 (Posto-linha=posto-coluna). $p = p^T$

Demonstração. Pelo teorema do núcleo e da imagem sabemos que $n+p = d$ e que $n^T+p^T = k$. Em particular, isto nos permite estabelecer as relações $n^T = k - p^T$ e $d - n = p$. Por outro lado, sabemos pelo Teorema 8.2.5 que

$$\dim N(f) + \dim N(f)^\perp = d \text{ e } \dim Im(f) + \dim Im(f)^\perp = k.$$

Logo, $\dim N(f)^\perp = d - n = p$ e $\dim Im(f)^\perp = k - p$.

Por outro lado, a Proposição 14.2.1 implica que $n^T \leq \dim Im(f)^\perp$ e que $p^T \leq \dim N(f)^\perp$. Como acabamos de ver que $\dim N(f)^\perp = p$, esta última igualdade nos diz que

$$p^T \leq p.$$

Além disso, como $\dim Im(f)^\perp = k - p$ e como $n^T = k - p^T$, a primeira desigualdade nos diz que

$$k - p^T \leq k - p \quad \Longrightarrow \quad p^T \geq p.$$

Portanto $p = p^T$. □

Capítulo 15

Equações Lineares

Esse é um capítulo fundamental dentro da estória que está sendo contada. Vamos ver como as funções lineares entre espaços vetoriais oferecem uma linguagem comum para se falar a respeito de uma multidão de problemas diferentes, desde problemas intrínsecos à álgebra linear até problemas oriundos de aplicações no mundo real.

A razão dessa ubiquidade das funções lineares é interessante. De fato, o objeto matemático mais conhecido por não-matemáticos são as equações. Todo mundo já tombou com alguma equação em um livro, museu, na TV, ou no YouTube. Existem equações *pop-art* como a conhecidíssima equação de Einsten $E = mc^2$. As equações fornecem uma ponte direta entre os objetos matemáticos abstratos, que só podem habitar a mente das mulheres e dos homens, e o mundo real. Uma equação numérica é uma igualdade que envolve *quantidades desconhecidas*, as *incógnitas* da equação e quantidades conhecidas, os *parâmetros* da equação. Uma *equação linear* é um conjunto de igualdades numéricas onde as incógnitas são números reais $x_1, ..., x_m$ e os parâmetros são números a_{ij}, com $i = 1, ..., n$ e $j = 1, ..., m$ e $b_1, ..., b_n$ (que supõe-se conhecidos nas aplicações) com a forma

$$\begin{array}{rcl} a_{11}x_1 + & \ldots & + a_{1m}x_m = b_1 \\ & \vdots & \\ a_{n1}x_1 + & \ldots & + a_{nm}x_m = b_n. \end{array}$$

As equações lineares modelam multidões de fenômenos, muitas vezes de forma precisa e outras vezes de forma aproximada e por isso desempenham um papel central em toda a

matemática e em todas as suas aplicações. E justamente, as equações lineares possuem uma relação profunda com as funções lineares, de modo que estudar uma coisa é quase a mesma coisa que estudar a outra, pois se você considerar a função linear $f : \mathbb{R}^m \to \mathbb{R}^n$ cuja matriz na base canônica é a matriz $A = [a_{ij}]_{n\times m}$ dos coeficientes da equação linear, $x = (x_1, ..., x_m) \in \mathbb{R}^m$ é o vetor incógnita e $b = (b_1, ..., b_n) \in \mathbb{R}^n$ é o vetor parâmetro da equação, então o conjunto de n igualdades numéricas escrito acima é a mesma coisa que a igualdade vetorial

$$f(x) = b.$$

Ou seja, os números $x_1, ..., x_m$ são soluções da equação com coeficientes a_{ij}, com $i = 1, ..., n$ e $j = 1, ..., m$ e $b_1, ..., b_n$ se, e somente se, o vetor $b = (b_1, ..., b_n)$ é a imagem de $x = (x_1, ..., x_m)$ pela função linear $f : \mathbb{R}^m \to \mathbb{R}^n$ cuja matriz na base canônica é $A = [a_{ij}]_{n\times m}$.

Através dessa conexão, ou dessa reinterpretação das equações lineares através das funções lineares, todo o entendimento sobre a geometria dessas funções entrará a serviço do entendimento das equações. Nesse capítulo vamos ver exemplos de equações e explorar consequências dessa conexão.

15.1 O problema do fluxo de tráfego

Suponhamos a seguinte situação, inspirada em fatos reais: com o intuito de melhor planejar as ações do *Programa Niterói de Bicicleta*, a prefeitura de Niterói decidiu fazer uma contagem de ciclistas nas principais vias da cidade. Para realizar esta tarefa, dedicou agentes da *Nitrans* para realizar a contagem de ciclistas que passam por cada via, em cada intervalo de 60 minutos. Como o número de agentes é limitado, a prefeitura decide contratar uma consultoria matemática que usará Álgebra Linear para melhor alocar os agentes. Na Figura 15.1 estão indicados quatro cruzamentos e a quantidade de ciclistas que passam por cada via, antes e depois de cada cruzamento. Observe que quatro destas quantidades não estão indicadas na figura, pois não é preciso dedicar agentes para fazer estas contagens, já que usando-se Álgebra Linear podemos determinar estas quantidades. Como a consultoria matemática pode ajudar a prefeitura de Niterói e calcular essas quantidades?

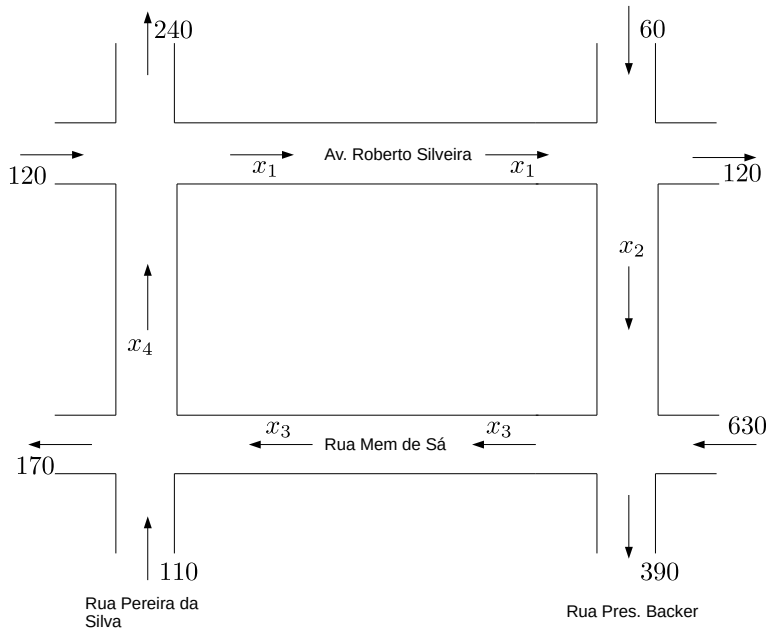

Figura 15.1: As quantidades de ciclistas x_1, x_2, x_3 e x_4 podem ser facilmente determinadas usando-se um pouco de Álgebra Linear.

Em primeiro lugar, vamos *modelar* o problema. Ou seja, baseados em observações empíricas vamos transformar o que vemos em equações matemáticas envolvendo as quantidades desconhecidas.

Observe que em cada cruzamento temos uma quantidade de ciclistas que entra e uma quantidade que sai. Para cada cruzamento, essas duas quantidades tem que ser iguais. Assim, olhando para o cruzamento da Av.Roberto Silveira com a Rua Pereira da Silva podemos traduzir a observação empírica na seguinte equação

$$120 + x_4 = x_1 + 240.$$

Analisando os três outros cruzamentos, obtemos as quatro equações

$$\begin{aligned} 120 + x_4 &= x_1 + 240 \\ 60 + x_1 &= 120 + x_2 \\ 630 + x_2 &= x_3 + 390 \\ 110 + x_3 &= 170 + x_4 \end{aligned}$$

Agora transformamos o problema concreto de determinar as quantidades de ciclistas que faltavam num problema matemático de resolver um sistema com quatro equações. Essa etapa é o que chamamos modelagem matemática do problema.

Mas e como resolvemos as equações?

De fato, esse sistema com quatro equações não é possível de ser resolvido de forma única, existem infinitas soluções para ele. Esse é um dos pontos cruciais do curso de Álgebra Linear: analisar quando um sistema possui solução, se é única. Veremos isso com detalhes no curso.

Para o nosso problema, isso quer dizer que as quatro equações não são suficientes para determinar a solução. Então, temos que voltar para a observação empírica e buscar mais informação (novas equações).

Olhando bem para a Figura 15.1 você vai ver que a quantidade total de ciclistas que entraram na região, é igual a quantidade total que saiu (920). Além disso, cada ciclista teve que passar no espaço entre dois cruzamentos, em alguma rua, e portanto foi computado por uma das quantidades x_1, x_2, x_3 ou x_4. Isso significa que a soma dessas quantidades deve coincidir com o total de ciclistas que passou pela região. Ou seja, temos uma nova equação

$$x_1 + x_2 + x_3 + x_4 = 920.$$

Agora podemos explicar como resolver o sistema com cinco equações. Observe que podemos as quatro primeiras da seguinte forma

$$\begin{aligned} x_1 &= -120 + x_4 \\ x_2 &= -60 + x_1 \\ x_3 &= 240 + x_2 \\ x_4 &= -60 + x_3 \end{aligned}$$

Além disso, podemos inserir a primeira equação na segunda, obtendo

$$x_2 = -60 - 120 - x_4 = -180 + x_4.$$

E podemos escrever a quarta equação na forma

$$x_3 = 60 + x_4.$$

Inserindo essas três equações na quarta equação obtemos

$$-120 + x_4 - 180 + x_4 + 60 + x_4 + x_4 = 920,$$

o que nos dá $4x_4 = 1160$, ou seja $x_4 = 290$. Com o valor de x_4 obtemos todos os demais.

15.2 Coeficientes da combinação linear

Vamos agora dar atenção a um problema intrínseco da álgebra linear. Suponha que o vetor $b \in \mathbb{R}^d$ seja combinação linear dos vetores $v^1, ..., v^k \in \mathbb{R}^d$. Ou seja, existem números reais $x_1, ..., x_k$ tais que

$$\sum_{\ell=1}^{k} x_\ell v^\ell = b. \tag{15.1}$$

Como podemos determinar os coeficientes $x_1, ..., x_k$ em função das coordenadas dos vetores $v^1, ..., v^k$ e b? Ora, considere a matriz $A = [a_{ij}]_{d\times k}$ cujas k colunas são as coordenadas de $v^1, ..., v^k$, respectivamente. Isto é

$$A = \begin{bmatrix} | & & | \\ v^1 & \dots & v^k \\ | & & | \end{bmatrix}.$$

Considere a função linear $f : \mathbb{R}^k \to \mathbb{R}^d$ cuja matriz na base canônica é A. O teorema a seguir estabelece a relação fundamental entre os três pilares da álgebra linear: combinações lineares, matrizes e equações lineares, sendo o coração dessa relação justamente o conceito de função linear.

Teorema 15.2.1. *Então, a igualdade vetorial $f(x) = b$ equivale a equação (15.1). Em outras palavras: a imagem do vetor $x \in \mathbb{R}^k$ pela função linear $f : \mathbb{R}^k \to \mathbb{R}^d$ é a combinação*

linear dos vetores coluna da matriz de f (na base canônica) cujos coeficientes são justamente as coordenadas de x.

Demonstração. Pela definição da matriz A temos que

$$v^j = \sum_{\ell=1}^{d} a_{\ell j} e_\ell = (a_{1j}, ..., a_{dj}) \in \mathbb{R}^d,$$

pois as coordenadas de v^j na base canônica fornecem a j-ésima coluna de A. Por outro lado, temos que

$$f(x) = \begin{bmatrix} \sum_{j=1}^{k} a_{1j} x_j \\ \vdots \\ \sum_{j=1}^{k} a_{dj} x_j \end{bmatrix} = \sum_{i=1}^{d} \left(\sum_{j=1}^{k} a_{ij} \right) e_i$$

Portanto,

$$\sum_{j=1}^{k} x_j v^j = \sum_{j=1}^{k} x_j \left(\sum_{i=1}^{d} a_{ij} e_i \right) = \sum_{j=1}^{k} \sum_{i=1}^{d} x_j a_{ij} e_i = \sum_{i=1}^{d} \left(\sum_{j=1}^{k} a_{ij} \right) e_i = f(x),$$

como queria demonstrar. □

O Teorema 15.2.1 diz, em particular, que determinar os coeficientes de uma combinação linear equivale a resolver uma equação linear. Além disso, podemos deduzir desse teorema um critério para dizer se quando uma equação linear admite solução ou não.

Corollário 15.2.2. *A equação*

$$\begin{aligned} a_{11}x_1 + &\quad ... \quad + a_{1m}x_m = b_1 \\ &\quad \vdots \\ a_{n1}x_1 + &\quad ... \quad + a_{nm}x_m = b_n. \end{aligned}$$

admite solução se, e somente se

$$b \in \mathrm{Im}(f),$$

onde $f : \mathbb{R}^m \to \mathbb{R}^n$ é a função linear cuja matriz na base canônica é a matriz $A = [a_{ij}]_{n \times m}$ dos coeficientes da equação linear. Em particular, se $\dim \mathrm{Im}(f) = n$ a equação sempre admite solução.

Finalmente, como uma pérola, tiramos uma consequência altamente não-trivial dessa conexão estabelecida entre equações e funções lineares. Podemos caracterizar completamente o conjunto de todas as soluções possíveis.

Corollário 15.2.3. *Seja $x \in \mathbb{R}^m$ uma solução da equação linear. Se $v \in N(f)$ é um elemento do núcleo de f então $x+v$ também é solução da equação. Em particular, somente três possibilidades existem: ou a equação não tem solução, ou tem uma única solução, ou possui infinitas soluções.*

Demonstração. Como $f(v) = 0$ se $f(x) = b$ então $f(x+v) = f(x) + f(v) = b$. A tricotomia segue pois se $\dim N(f) > 0$ então $N(f)$ possui infinitos elementos. □

15.3 O Método da eliminação de Gauss

15.3.1 Solução de equações lineares

Considere a matriz

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix},$$

e o vetor $y = (10, 11, 12)$. Vamos resolver o problema de se encontrar $x = (x_1, x_2, x_3) \in \mathbb{R}^3$ de forma que $A(x) = y$ usando um método que funciona de forma geral para resolver esse tipo de problema. A ideia é que se consideramos o sistema de equações

$$\begin{aligned} x_1 + 2x_2 + 3x_3 &= 10 \\ 4x_1 + 5x_3 + 6x_4 &= 11 \\ 7x_1 + 8x_2 + 9x_3 &= 12 \end{aligned}$$

que é equivalente à igualdade vetorial $A(x) = y$, então podemos se multiplicamos cada equação por número e somamos obtemos um novo sistema cujas incógnitas são as mesmas. Ou seja, podemos cada linha do sistema por combinações lineares das linhas, obtendo assim um novo sistema cuja solução é a mesma do sistema original. A ideia é fazer isso de forma sistemática até simplificar a equação de forma que a solução seja evidente.

$$\begin{bmatrix} 1 & 2 & 3 & 10 \\ 4 & 5 & 6 & 11 \\ 7 & 8 & 9 & 12 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 3 & 10 \\ 0 & 3 & 6 & 29 \\ 0 & 6 & 12 & 58 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 3 & 10 \\ 0 & 3 & 6 & 29 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{15.2}$$

Concluímos assim, que a equação procurada é equivalente ao sistema

$$\begin{aligned} x_1 + 2x_2 + 3x_3 &= 10 \\ 3x_2 + 6x_3 &= 29 \end{aligned} \tag{15.3}$$

Vemos que qualquer vetor $(x_1, x_2, x_3) \in \mathbb{R}^3$ cujas coordenadas cumprem as relações

$$\begin{aligned} x_3 &= \frac{29}{6} - \frac{x_2}{2} \\ x_1 &= \frac{-9}{2} + \frac{x_2}{2} \end{aligned}$$

é solução do sistema.

15.3.2 Dimensão do subespaço gerado

Considera agora o seguinte problema: os vetores $x = (1, 2, 3)$, $y = (4, 5, 6)$ e $z = (7, 8, 9)$ do $\mathbb{R}^3$ são LI ou LD? E qual a dimensão do subespaço gerado por $\{x, y, z\}$? Vamos ver como o método de "eliminação Gaussiana"que usamos acima em (15.2) responde ambas as perguntas.

Com efeito, se a gente ignora a quarta de coluna de cada matriz, os vetores em questão são as linhas da primeira matriz. Ou seja, considere a matriz

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix},$$

cujas linhas são $\ell_1 = x$, $\ell_2 = y$ e $\ell_3 = z$. A segunda matriz em (15.2),

$$B = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 3 & 6 \\ 0 & 6 & 12 \end{bmatrix}$$

foi obtida a partir de A colocando na segunda e na terceira linhas, respectivamente, os vetorres

$$4\ell_1 - \ell_2 = 4x - y \text{ e } 7\ell_1 - \ell_2 = 7x - z.$$

Dessa forma, vemos que

Em cada etapa do método de eliminação da Gauss, as linhas da próxima matriz vão ser combinações lineares das linhas da matriz atual. Além disso, as combinações lineares são escolhidas de forma a tornar a matriz triangular superior.[1]

Depois, para formar a terceira matriz a partir da segunda trocamos a terceira linha por uma combinação linear que faz -2 vezes a segunda linha de B somado com a terceira linha de B. Nesse momento obtemos o vetor nulo. A consequência disso é que existe portanto uma combinação linear não-trivial (nem todos os coeficientes são zero) dos vetores iniciais x, y, z que resulta no vetor nulo. Por isso deduzimos que x, y, z são LD. Em particular, isso prova que a dimensão do subespaço gerado é menor do 3, porque ela pode ser no máximo três, porém como os vetores são LD ela não pode ser exatamente 3 já que um deles tem que ser combinação linar dos demais.

E qual a dimensão do subespaço gerado? Ora, observe que

$$4x - y = (0, 3, 6) \in S(\{x, y, z\})$$

e que os vetores $x = (1, 2, 3)$ e $(0, 3, 6)$ são LI, pois não são colineares. Essa é a razão de colocar as entradas abaixo da diagonal igual a zero, porque é muito fácil ver se os vetores obtidos são LI: todas as linhas que tiverem na entrada diagonal um número não-nulo, vão ser LI porque vão ter zeros em posições diferentes. Assim exibimos dois vetores no subespaço gerado $S(\{x, y, z\})$ que são L, x e $4x - y$. Logo a dimensão do subespaço gerado é 2.

15.3.3 Cálculo do posto e da nulidade de uma matriz

Vamos agora ver como podemos usar o método da eliminação de Gauss para determinar o posto e a nulidade de uma matriz. Antes de iniciar a descrição do algorítimo vamos ver

[1] Uma matriz é triangular superior se todas as entradas abaixo da diagonal são nulas.

alguns conceitos.

Definição 15.3.1. *Sejam X e Y espaços vetoriais de dimensão finita e seja $f : X \to Y$ uma função linear. A* nulidade *de f é a dimensão de seu núcleo. O* posto *de f é a dimensão da sua imagem.*

Com essa definição o teorema do Núcleo e da Imagem diz que

$$\textbf{posto} + \textbf{nulidade} = \dim X$$

A pergunta que queremos responder aqui é bem simples: dada $f : X \to Y$, como calcular o posto e a nulidade de f? Pela fórmula do núcleo e da imagem acima, já sabemos que basta calcular uma dessas quantidades.

Para resolver esse problema usando o método de eliminação, vamos considerar bases em X e Y. Então seja $\mathcal{B} = \{b^1, ..., b^d\}$ uma base de X e $\mathcal{A} = \{y^1, ..., y^k\}$ uma base de Y. Para cada $j = 1, ..., d$ sejam a_{ij} as respectivas coordenadas de $f(b^j)$ na base $\mathcal{A}$ de Y. Assim,

$$f(b^j) = \sum_{i=1}^{k} a_{ij} y^i, \quad \forall\ j = 1, ..., d.$$

Como vimos nos capítulos anteriores isso permite representar f por uma matriz $A = [a_{ij}]_{k \times d}$. As colunas dessa matriz são as coordenadas de $f(b^j)$ na base $\mathcal{A}$ de Y. Se $y \in Im(f)$ é um vetor arbitrário na imagem de f então existe $x \in X$ tal que $y = f(x)$. Como $\mathcal{B}$ é base X devem existir números reais $x_1, ..., x_d$ tais que

$$x = \sum_{j=1}^{d} x_j b^j.$$

Como f é linear isso implica que

$$f(x) = \sum_{j=1}^{d} x_j f(b^j).$$

Isso prova o seguinte resultado:

Lema 15.3.2. *A imagem de f é o subespaço gerado velos vetores coluna de sua matriz, ou seja*

$$Im(f) = \mathcal{S}(\{f(b^1), ..., f(b^j)\}).$$

Exemplo 15.3.3. *Qual o posto da matriz* $A = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}$. *Lembrando que as colunas de A são as imagens dos vetores da base canônica, de acordo com o lema 15.3.2 devemos calcular a dimensão do subespaço gerado pelos vetores* $A(e_1) = (1, 2, 3)$, $A(e_2) = (4, 5, 6)$ *e* $A(e_3) = (7, 8, 9)$. *Isso implica aplicar o método de eliminação à matriz*

$$A^T = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix},$$

que é a transposta de A. Já vimos acima que a dimensão do subespaço gerado pelas linhas dessa matriz é 2, portanto o posto de A é dois.[2]

15.4 Interpolação polinomial

Aqui vamos mostrar uma aplicação muito interessante das equações lineares a um problema de matemática, fora da álgebra linear, mas que possui muita utilidade no mundo real. A questão é a seguinte: como encontrar um polinômio cujo gráfico passe por alguns pontos dados do plano?

Esse problema aparece em situações onde parece haver uma função que governa um determinado fenômeno e se conhece apenas um número finito de medições. O polinômio interpolador serve para fazer previsões.

Então, procuramos um polinômio de grau d

$$p(t) = \sum_{\ell=0}^{n} a_\ell t^\ell$$

que satisfaça $p(t_i) = y_i$ para $i = 1, ...n$. Ou seja, cujo gráfico passe pelos pontos (t_i, y_i) no

[2]Um resultado muito útil mas que não vamos explorar muito é que o posto de uma matriz é igual ao posto da sua transposta, portanto tanto faz fazer a eliminação com as linhas ou com as colunas de A.

plano. Essas n igualdades podem ser reescritas usando a *Matriz de Vandermonde:*

$$V = \begin{bmatrix} 1 & t_1 & \dots & t_1^{d-1} & t_1^d \\ 1 & t_2 & \dots & t_2^{d-1} & t_2^d \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 1 & t_n & \dots & t_n^{d-1} & t_n^d \end{bmatrix}.$$

Assim, a condição de que os coeficientes $a_0, ..., a_d$ satisfaçam as equações de interpolação equivale a equação vetorial

$$V(a) = y,$$

onde $a = (a_0, ..., a_d) \in \mathbb{R}^{d+1}$ e $y = (y_1, ..., y_n) \in \mathbb{R}^n$.

15.4.1 Implementação em Julia

A matriz de Vandermonde permite obter de forma algorítmica, e muito simples, o polinômio interpolador. Por exemplo, considere o seguinte problema:

Problema 3. *Encontre um polinômio de grau 3 que satisfaça* $p(1) = 1$*,* $p(2) = 8$ *e* $p(3) = 27$*.*

Para resolver esse problema, usando a Matriz de Vandermonde, em Julia, primeiro escrevemos a matriz de Vandermonde associada ao vetor $t = (1, 2, 3) \in \mathbb{R}^3$. De fato, podemos escrever uma função em Julia que recebe o grau do polinômio interpolador e a abcissa dos dados e devolve diretamente a matriz de Vandermonde:

```
function vand(t,d)
n=length(t)
V=zeros(n,d+1)
for i=1:n
for j=1:d+1
V[i,j]=t[i]^(j-1)
end
end
return V
end
```

No nosso caso, $t = (1, 2, 3)$ e $d = 3$. Declaramos isso e Julia nos devolve a matriz de Vandermonde:

```
t=[1,2,3]
d=3
vand(t,d)
```

E Julia nos devolve a matriz

$$V = \begin{bmatrix} 1.0 & 1.0 & 1.0 & 1.0 \\ 1.0 & 2.0 & 4.0 & 8.0 \\ 1.0 & 3.0 & 9.0 & 27.0 \end{bmatrix}$$

Agora a gente define uma função que retorna a solução da equação $V(a) = y$:

```
function interpol(t,y,d)
n=length(y)
x=vand(t,d)\ y
return x
end
```

Com os parâmetros dados, Julia nos dá o polinômio

$$0.030927835051541036 - 0.0567010309278361x + 0.03092783505154996x^2 + 0.9948453608247413x^3$$

Para ter essa resposta, precisamos instalar o pacote Polynomials:

```
import Pkg
Pkg.add("Polynomials")
using Polynomials
```

Finalmente, podemos visualizar a resposta usando PyPlots:

```
using Pkg Pkg.add("PyPlot") using PyPlot p = Poly(interpol(t,y,d))
xs = range(-1,stop=3,length=100); ys = map(x -¿ polyval(p, x), xs)
plot(xs, ys, color=:blue, linewidth=2.0, leg=:false)
plot!(t,y,color=:red,marker=:circle,leg=:false)
```

O que fornece a resposta:

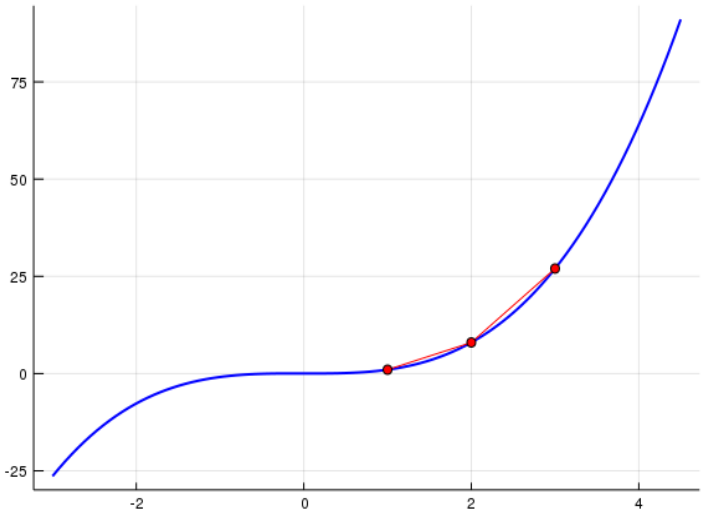

Figura 15.2: Os pontos vermelhos são os pontos a interpolar, e o gráfico azul é o polinômio interpolador.

Capítulo 16

Sistemas Dinâmicos Lineares

Capítulo 17

Quociente e Dualidade

Esse capítulo é destinado aos amantes da matemática.

Nessa seção vamos estudar uma ferramenta abstrata para contagem de dimensão. Tal ferramenta está fundamentada numa ideia muito simples: descartar informação irrelevante. Para fazer esse descarte, "passamos ao quociente"para que reste apenas a parte do espaço que é relevante. Esse processo pode ser feito de muitas formas diferentes, dependendo do que queiramos descartar.

Relações de equivalência

Muitas vezes os elementos de um conjunto podem ser identificados entre si por uma característica em comum. Essa é uma ideia corriqueira, como identificar pessoas num grupo que usam camisas da mesma cor.

A abstração dessa ideia é o que chamamos em matemática de uma *relação de equivalência*.

Definição 17.0.1. *Seja X um conjunto. Uma* relação *em X é um subconjunto $\mathcal{R} \subset X \times X$ do produto cartesiano. Se $(x, y) \in \mathcal{R}$ denotamos $x \sim y$.*

Por simplicidade de notação é costume indicar uma relação apenas especificando o que é necessário o par x, y satisfazer para que estejam relacionados.

Exemplo 17.0.2. *No conjunto $\{1, 2, 3, 4\} \subset \mathbb{N}$, podemos definir a relação $x \sim y$ se e somente*

se $x > y$. Nesse caso, o subconjunto $\mathcal{R}$ do produto cartesiano para essa relação será

$$\{(2,1),(3,2),(3,1),(4,3),(4,2),(4,1)\}.$$

Definição 17.0.3. *Uma relação num conjunto X é dita uma* relação de equivalência *se satisfaz as condições a seguir*

1. $x \sim x$, $\forall\, x \in X$ (reflexividade)
2. $x \sim y \iff y \sim x$ (simetria)
3. $x \sim y$, $y \sim z \implies x \sim z$ (transitividade)

Uma relação de equivalência num conjunto X permite identificar elementos que possuam alguma característica em comum e estudar o espaço X somente ao olhos dessa característica. Por exemplo, se você quer contar contas cores de olhos há num grupo de pessoas, você vai fazer um processo semelhante mentalmente: agrupas as pessoas com olhos de mesma cor e contar quantos grupos você formou.

Definição 17.0.4. *Seja X um conjunto munido de uma relação de equivalência. Dado $x \in X$, a* classe de equivalência *de x, denotada por $[x]$ é o conjunto de todos os elementos $y \in X$ que são relacionados com x. Em outras palavras*

$$[x] \stackrel{\text{def}}{=} \{y \in X;\ y \sim x\}.$$

Assim como uma pessoa não pode ter olhos com duas cores, um ponto não pode pertencer a duas classes de equivalência diferentes.

Proposição 17.0.5. *Seja X um conjunto munido de uma relação de equivalência e sejam $x, y \in X$. Então,*

$$[x] \cap [y] \neq \emptyset \implies [x] = [y].$$

Demonstração. Suponhamos que exista $z \in [x] \cap [y]$. Então, por definição $z \sim x$ e $z \sim y$. Pela transitividade isso implica que $x \sim y$, e pela simetria $y \sim x$. Agora, se $v \in [x]$ (ou seja $v \sim x$) então, como $x \sim y$, por transitividade segue que $v \sim y$, ou seja $v \in [y]$. Isso prova que $[x] \subset [y]$. Com um argumento análogo podemos provar que $[y] \subset [x]$. Portanto $[x] = [y]$, como queria demonstrar. □

A Proposição 17.0.5 diz que duas classes de equivalência ou são iguais ou são disjuntas. Além disso, toda classe de equivalência é não vazia, já que todo elemento está relacionado com sigo mesmo. Desse modo classes de equivalência produzem *partições* do conjunto X: uma maneira de dividir X em "regiões"(subconjuntos) dois a dois disjuntos cuja união é todo o X.

Com um pouco mais de formalidade, considere $\mathcal{P}(X)$ o conjunto das partes de X, i.e. o conjunto cujos elementos são os subconjuntos de X. Uma relação de equivalência em X determina uma função

$$\begin{aligned} \pi : \quad X &\to \mathcal{P}(X) \\ x &\mapsto \pi(x) = [x], \end{aligned}$$

chamada a aplicação quociente. O conjunto quociente, ou simplesmente o *quociente* de X pela relação de equivalência é a imagem da aplicação quociente. Ou seja,

$$X/\sim \stackrel{\text{def}}{=} \pi(X).$$

Em termos menos formais, o quociente de X pela relação de equivalência é o agregado de todas as classes de equivalência.

Quociente de um espaço por um subespaço

Suponhamos agora que X seja um espaço vetorial e $F \subset X$ seja um subespaço.

Definição 17.0.6. *Dois elementos $x, y \in X$ são ditos* congruentes módulo F *se $x - y \in F$.*

Isso permite introduzir uma relação em X, dizendo

$$x \sim y \iff x - y \in F$$

Proposição 17.0.7. *A relação definida acima é uma relação de equivalência.*

Demonstração. Queremos provar que a relação é reflexiva, simétrica e transitiva.

Reflexividade Seja $x \in X$ arbitrário. Como F é um subespaço, $0 \in F$. Como $x - x = 0$, concluímos que $x \sim x$.

Simetria Sejam $x, y \in X$, com $x \sim y$. Isso significa, por definição, que $x - y \in F$. Como F é um subespaço, $-(x-y) \in F$, logo $y - x \in F$. Ou seja, $y \sim x$.

Transitividade Sejam $x, y, z \in X$. Suponhamos que $x \sim y$ e que $y \sim z$. Então, $x - y \in F$ e $y - z \in F$. Como F é subespaço,

$$\begin{aligned} (x-y)+(y-z) &\in F \\ \iff x - z &\in F. \end{aligned}$$

Portanto $x \sim z$. Isso prova que a relação é transitiva. □

Podemos considerar o quociente $X/\sim$ de X pela relação de equivalência congruência módulo subespaço. No caso particular dessa relação é costumeiro indicar $X/F \stackrel{\text{def}}{=} X/\sim$.

Observe que nesse caso começamos com um conjunto que possui uma certa estrutura (é um espaço vetorial) e tomamos o quociente pela relação congruência módulo subespaço, obtendo, como em geral, um *novo conjunto* X/F. Uma pergunta natural é se esse novo conjunto pode herdar a estrutura de espaço vetorial.

Para isso precisamos dizer como "somar"subconjuntos de X e como multiplicá-los por um escalar. Uma proposta é a seguinte:

$$\begin{aligned} [x]+[y] &\stackrel{\text{def}}{=} [x+y] \\ \lambda[x] &\stackrel{\text{def}}{=} [\lambda x],\ \forall\, \lambda \in \mathbb{R}. \end{aligned}$$

Um problema com essa definição é o seguinte: definimos a soma de duas classes como a classe da soma. No entanto, uma classe contém vários pontos e, por exemplo, podemos ter $[x] = [z]$ e $[y] = [v]$. Para que a definição não seja ambígua devemos ter $[x+y] = [z+w]$. É isso que vamos ver agora.

Lema 17.0.8. *Se $z \in [x]$, $v \in [y]$ então $[z+v] = [x+y]$ e $[\lambda x] = [\lambda z]$.*

Demonstração. Começamos provando que $[z+v] = [x+y]$. Seja $u \in [z+v]$. Então, $u \sim z+v$, ou seja $u - (z+v) \in F$. Por hipótese, sabemos que $x - z \in F$ e $y - v \in F$. Portanto, como F é subespaço

$$\begin{aligned} u-(z+v)-(x-z)-(y-v) &\in F \\ \iff u-(x+y) &\in F, \end{aligned}$$

e portanto $u \in [x+y]$. Isso prova que $[z+v] \subset [x+y]$. Com um argumento análogo, provamos que $[x+y] \subset [z+v]$. Isso mostra que $[x+y] = [z+v]$, como queria. Vamos agora demonstrar que $[\lambda x] = [\lambda z]$. Seja $u \in [\lambda x]$. Então, $u - \lambda x \in F$. Como $x - z \in F$ temos $\lambda x - \lambda z \in F$, pois F é subespaço. Logo,

$$\begin{aligned} u - \lambda x + (\lambda x - \lambda z) \quad &\in \quad F \\ \iff u - \lambda z &\in F, \end{aligned}$$

e portanto $u \sim \lambda z$, ou seja $u \in [\lambda z]$. Com um argumento análogo provamos que $[\lambda z] \subset [\lambda x]$. A prova do lema está completa. □

Proposição 17.0.9. *As operações definidas acima tornam X/F um espaço vetorial.*

Demonstração. Devemos verificar que os axiomas de espaço vetorial são satisfeitos. Por exemplo, vamos verificar que a soma é comutativa. Para isso, usamos que a soma em X é comutativa, já que X é um espaço vetorial. Ela garante que $x+y = y+x$ e isso implica que $[x+y] = [y+x]$. Como consequência obtemos

$$[x] + [y] = [x+y] = [y+x] = [y] + [x].$$

Vamos verificar a distributividade. Sejam $x, y \in X$ e $\lambda \in \mathbb{R}$. Como $\lambda(x+y) = \lambda x + \lambda y$, temos que $[\lambda(x+y)] = [\lambda x + \lambda y]$. Com isso obtemos,

$$\lambda([x] + [y]) = \lambda[x+y] = [\lambda(x+y)] = [\lambda x + \lambda y] = [\lambda x] + [\lambda y].$$

Para mostrar que X/F possui um vetor nulo, observe que se $x \in F$ então $[x] = [0]$, pois $x - 0 \in F$, duas classes de equivalência são sempre iguais ou disjuntas. Assim, temos que

$$[0] + [x] = [0 + x] = [x].$$

As demais demonstrações são análogas. □

Exemplo 17.0.10. *Seja $X = \mathbb{R}^2$ e $F = \{(0, x); x \in \mathbb{R}\}$. Então, como é fácil ver, F é um subespaço de $\mathbb{R}^2$. Então a relação de congruência módulo F pode ser traduzida como $x \sim y$ se, e só se, x e y possuem a mesma coordenada na primeira entrada.*

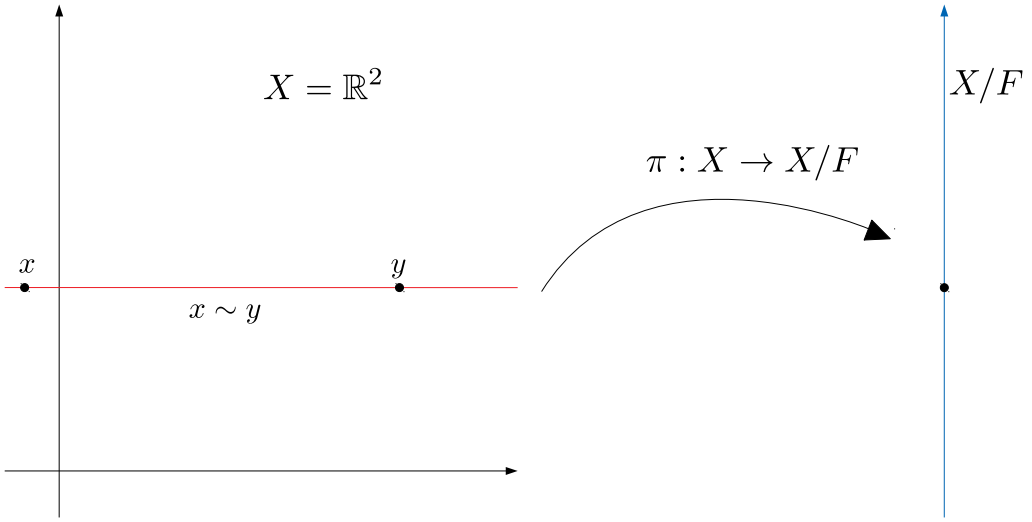

Figura 17.1: Ao formar o quociente descartamos a primeira coordenada

A passagem ao quociente é uma ferramente poderosa para contagem de dimensão. Isso é o que vemos no teorema abaixo, no qual um princípio fundamental da natureza pode ser visto em funcionamento: a soma da informação descartada com a informação restante dá o total de informação que tínhamos antes do descarte.

Teorema 17.0.11. *Seja X um espaço vetorial de dimensão finita e $F \subset X$ um subespaço. Então*

$$\dim F + \dim X/F = \dim X.$$

Demonstração. Seja $\mathcal{F} = \{v_1, .., v_n\}$ uma base de F. Em particular, $\mathcal{F}$ é um conjunto LI em X. Aplicando o Teorema 7.4.8, podemos completar o conjunto $\mathcal{F}$ de modo a obter uma base de X. Portanto, existem vetores $\{x_1, ..., x_k\} \stackrel{\text{def}}{=} \mathcal{B}$ tais que $\mathcal{F} \cup \mathcal{B}$ é uma base de X. Em particular, como $\dim F = n$, temos que

$$d = \dim X = n + k = \dim F + k.$$

Logo, o teorema estará provado se provarmos que $k = \dim X/F$. Isso, por sua vez, será atingido se provarmos que $\pi(\mathcal{B}) = \{[x_1], ..., [x_k]\}$ é uma base de X/F.

Vamos começar provando que $\pi(\mathcal{B})$ é LI em X/F. Suponhamos que existem números reais $\alpha_1, ..., \alpha_k$ tais que

$$\sum_{l=1}^{k} \alpha_l [x_l] = 0.$$

Pela definição das operações de X/F e pelos axiomas de espaço vetorial temos que $\sum_{l=1}^{k} \alpha_l [x_l] = [\sum_{l=1}^{k} x_l]$. Como o vetor nulo de X/F é $0 = [0] = \pi(F)$, concluímos que

$$\sum_{l=1}^{k} \alpha_l x_l \in F.$$

Portanto existe $y \in F$ tal que $\sum_{l=1}^{k} \alpha_l x_l = y$. Como $\mathcal{F}$ é uma base de F, existem números reais $\beta_1, ..., \beta_n$ tais que $y = \sum_{l=1}^{n} \beta_l v_l$. Então,

$$\sum_{l=1}^{k} \alpha_l x_l - \sum_{l=1}^{n} \beta_l v_l = 0.$$

Como $\mathcal{F} \cup \mathcal{B}$ é uma base de X, em particular é um conjunto LI, segue que todos os coeficientes α_l (e β_l) são nulos. Pelo Corolário 7.3.5, segue que $\pi(\mathcal{B})$ é LI.

Em seguida, vejamos que é uma base de X/F. Seja $[x] \in X/F$ um elemento arbitrário. Temos que x pode ser escrito como combinação linear dos elementos da base $\mathcal{F} \cup \mathcal{B}$. Ou seja, dado $x \in X$ existem números reais $\alpha_1, ..., \alpha_k$ e $\beta_1, ..., \beta_n$ tais que

$$x = \sum_{l=1}^{n} \beta_l v_l + \sum_{l=1}^{k} \alpha_l x_l = v + y,$$

onde $v \stackrel{\text{def}}{=} \sum_{l=1}^{n} \beta_l v_l$ e $y \stackrel{\text{def}}{=} \sum_{l=1}^{k} \alpha_l x_l$. Observe que $v \in F$. Isso implica que $x - y = v \in F$ e portanto, $x \sim y$. Logo, $[x] = [y]$ e como $[v] = [0]$ concluímos que

$$[x] = [\sum_{l=1}^{k} \alpha_l x_l] = \sum_{l=1}^{k} \alpha_l [x_l].$$

Provamos assim que $[x]$ pode ser escrita como combinação linear dos vetores $[x_l]$, $l = 1, ..., k$. Isso estabelece que $\pi(\mathcal{B})$ é gerador de X/F e completa a prova. □

Vamos terminar esse parágrafo com um lema técnico sobre a operação de passagem ao quociente. Ele será útil nas aplicações.

Lema 17.0.12. *Sejam $F \subset E \subset X$ subespaços e X/F o quociente com $\pi : X \to X/F$ a aplicação quociente. Então $\pi(E) \subset X/E$ é um subespaço e* $\dim \pi(E) = \dim E - \dim F$.

Demonstração. Para ver que $\pi(E)$ é um subespaço, observe que $\pi(0) = [0] \in \pi(E)$, pois $0 \in E$. Sejam $\pi(x), \pi(y) \in \pi(E)$. Observe que, pela definição de π e das operações em X/F, temos que, qualquer que seja $\lambda \in \mathbb{R}$, temos

$$\pi(x) + \lambda\pi(y) = [x] + \lambda[y] = [x + \lambda y] = \pi(x + \lambda y).$$

Por isso, $\pi(x) + \pi(y) = \pi(x + y) \in E$ e $\lambda\pi(x) \in \pi(E)$. Logo, $\pi(E)$ é de fato um subespaço.

Por outro lado, como $F \subset E$ é um subespaço, podemos considerar o quociente E/F. Nesse caso, a relação de equivalência será a mesma que antes: $x, y \in E$ são relacionados $\iff x - y \in F$. Isso mostra que $\pi(E) = E/F$.

Aplicando o Teorema 17.0.11 concluímos que $\dim \pi(E) + \dim F = \dim E$, o que dá a formula procurada. □

Soma de Subespaços

Vamos agora aplicar a nossa ferramenta de contagem de dimensão para demonstrar uma fórmula clássica da álgebra linear: a dimensão da soma de dois subespaços.

Definição 17.0.13. *Seja X um espaço vetorial e sejam $E, F \subset X$ dois subespaços. A* soma *deles é o conjunto*

$$E + F \stackrel{\text{def}}{=} \{v \in X; \exists x \in E,\ y \in F;\ v = x + y\}$$

formado por todos os vetores que podem ser escritos como soma de um vetor em E com um vetor em F.

O conjunto $E + F \subset X$ é um subespaço de X. Como $0 \in E \cap F$, claramente $0 \in E + F$. Além disso, se $v, v' \in E + F$, existem $x, x' \in E$ e $y, y' \in F$ tais que $v = x + y$ e $v' = x' + y'$. Assim,

$$v + v' = (x + x') + (y + y') \in E + F.$$

Similarmente, vemos que $\lambda v \in E + F$, para todo $\lambda \in \mathbb{R}$. Como calcular a dimensão de $E + F$ a partir das dimensões de E e de F?

Exemplo 17.0.14.

1. *Sejam $F = \mathcal{S}(\{e_1, e_2\}) \subset \mathbb{R}^3$ e $E = \mathcal{S}(\{e_1, e_3\})$. Então, podemos ver que $E + F = \mathbb{R}^3$, e $\dim E = \dim F = 2$.*

2. *Sejam agora $E = \mathcal{S}(\{e_1, e_2\}) \subset \mathbb{R}^4$ e $F = \mathcal{S}(\{e_3, e_4\}) \subset \mathbb{R}^4$. Então, $E + F = \mathbb{R}^4$ e $\dim E = \dim F = 4$. Observe que $E \cap F = \{0\}$: a interseção entre eles é trivial.*

Lembre que a interseção entre dois subespaços é sempre um subespaço. O caso em que essa interseção é o subespaço trivial que consiste somente no vetor nulo é especial.

Definição 17.0.15. *Se $E \cap F = \{0\}$ denotamos a soma por $E \oplus F$ e dizemos que $Y = E \oplus F$ é a* soma direta de E e F.

Proposição 17.0.16. *Se $Y = E \oplus F$ então todo vetor em Y pode ser escrito de modo único como soma de um vetor em E com um vetor em F. Além disso,*

$$\dim Y = \dim E + \dim F.$$

Demonstração. Pela definição de soma, todo vetor em Y se escreve como soma de um vetor em E com um vetor em F. Suponhamos que para um certo $y \in Y$ isso possa ser feito de dois modos: existem $v, v' \in E$ e $u, u' \in F$ tais que

$$y = v + u = v' + u'.$$

Então

$$v - v' = u' - u.$$

Como $v - v' \in E$ e $u - u' \in F$ e como $E \cap F = \{0\}$, concluímos que $v = v'$ e $u = u'$, portanto na verdade só existe uma forma.

Agora, observe que $E \subset Y$, pois se $x \in E$ então $x + 0 \in E \oplus F$. Similarmente, $F \subset Y$. Tome $\{v_1, ..., v_k\}$ uma base de E e $\{u_1, ..., u_n\}$ uma base de F. Em particular, $\dim E = k$ e $\dim F = n$. Vamos provar que $\{v_1, ..., v_k, u_1, ..., u_n\}$ é uma base de Y. Se $y \in Y$ existem $v \in E$ e $u \in F$, únicos, tais que $y = v + u$. Como $v \in E$ e $u \in F$ existem $\alpha_1, ..., \alpha_k$ e $\beta_1, ..., \beta_n$ números reais tais que $v = \sum_{l=1}^{k} \alpha_l v_l$ e $u = \sum_{l=1}^{n} \beta_l u_l$ e então, $u + v$ se escreve como combinação linear dos vetores $\{v_1, ..., v_k, u_1, ..., u_n\}$. Isso prova que o conjunto é gerador. Suponha agora que

$$\sum_{l=1}^{k} \alpha_l v_l + \sum_{l=1}^{n} \beta_l u_l = 0.$$

Então $\sum_{l=1}^{k} \alpha_l v_l = -\sum_{l=1}^{n} \beta_l u_l$ pertencem ambos a $E \cap F = \{0\}$. Como $\{v_1, ..., v_k\}$ e $\{u_1, ..., u_n\}$ são ambos conjuntos LI, concluímos que todos os α's e todos os β's são nulos. Isso prova que $\{v_1, ..., v_k, u_1, ..., u_n\}$ é LI. Portanto, $\dim Y = k + n = \dim E + \dim F$. □

E quando a interseção $E \cap F$ possui dimensão positiva, como calcular a dimensão da soma $E + F$? Aqui vamos usar a ideia de passar ao quociente para reduzir esse problema ao caso que já sabemos resolver.

Teorema 17.0.17. *Sejam X um espaço vetorial de dimensão finita e $E, F \subset X$ dois subespaços. Então,*

$$\dim E + F = \dim E + \dim F - \dim E \cap F.$$

Demonstração. Seja $Y = E + F$. Então, como antes, E e F são subespaços de Y, assim como $Y_0 = E \cap F$ é subespaço de Y. Considere o quociente Y/Y_0, e a aplicação quociente associada $\pi : Y \to Y/Y_0$. Nesse caso, temos duas cadeias de subespaços: $Y_0 \subset E \subset Y$ e $Y_0 \subset F \subset Y$. Aplicando o Lema 17.0.12 vemos que $\pi(E)$ e $\pi(F)$ são ambos subespaços de Y/Y_0 e $\dim \pi(E) = \dim E - \dim Y_0$ e $\dim \pi(F) = \dim F - \dim Y_0$.

Além disso, se $[p] \in \pi(E) \cap \pi(F)$ então $p \in E \cap F$. Com efeito, a hipótese $[p] \in \pi(E) \cap \pi(F)$ implica que existe $x \sim p$ tal que $x \in E \cap F$. Mas $x - p \in Y_0$ e $x \in Y_0$ implicam $p \in Y_0$, como afirmado. Como $p \in Y_0$ implica $[p] = [0]$ isso prova que

$$\pi(E) \cap \pi(F) = [0].$$

Aplicando a Proposição 17.0.16 e o Teorema 17.0.11 temos que

$$\begin{aligned} \dim Y/Y_0 &= \dim \pi(E) + \dim \pi(F) \\ \iff \dim Y - \dim Y_0 &= \dim E - \dim Y_0 + \dim F - \dim Y_0 \\ \iff \dim Y &= \dim E + \dim F - \dim Y_0, \end{aligned}$$

que é a fórmula anunciada. □

Exercício 1. *Prove que se um espaço vetorial E é a soma direta de subespaços $E_1, ..., E_k$ então*

$$\dim E = \sum_{l=1}^{k} \dim E_l.$$

Exercício 2. *Prove que se $F \subset E$ é um subespaço de E tal que $\dim F = \dim E$ então $F = E$.*

Capítulo 18

Produto de matrizes

Sejam $A : \mathbb{R}^d \to \mathbb{R}^k$ e $B : \mathbb{R}^n \to \mathbb{R}^d$ funções lineares representadas nas bases canônicas pelas matrizes

$$A = \begin{bmatrix} a_{11} & \dots & a_{1d} \\ \vdots & \dots & \vdots \\ a_{k1} & \dots & a_{kd} \end{bmatrix} \quad \text{e} \quad B = \begin{bmatrix} b_{11} & \dots & b_{1n} \\ \vdots & \dots & \vdots \\ b_{d1} & \dots & b_{dn} \end{bmatrix}$$

Observe que as linhas de A são os vetores

$$\ell_1^A = (a_{11}, ..., a_{1d}) \in \mathbb{R}^d, ..., \ell_k^A = (a_{k1}, ..., a_{kd}) \in \mathbb{R}^d,$$

ao passo que as colunas de B são os vetores

$$c_1^B = (b_{11}, ..., b_{d1}) \in \mathbb{R}^d, ..., c_n^B = (b_{1n}, ..., b_{dn}) \in \mathbb{R}^d.$$

Definição 18.0.1. *A matriz* produto de A por B *é a matriz definida por*

$$AB \stackrel{\text{def}}{=} \begin{bmatrix} \langle \ell_1^A, c_1^B \rangle & \dots & \langle \ell_1^A, c_1^B \rangle \\ \vdots & \dots & \vdots \\ \langle \ell_k^A, c_1^B \rangle & \dots & \langle \ell_k^A, c_n^B \rangle \end{bmatrix}$$

Exemplo 18.0.2. *Sejam* $A = \begin{bmatrix} 2 & 0 \\ 0 & 1/3 \end{bmatrix}$ *e* $B = \begin{bmatrix} 1/2 & 0 \\ 0 & 3 \end{bmatrix}$. *Então* $AB = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

Exemplo 18.0.3. *Vamos usar* ***Julia*** *para efetuar o cálculo do produto de*

$$A = \begin{bmatrix} 11 & 34 & 6.7 & 4.9 \\ 1 & 1 & 1 & 1 \\ 2 & 4 & 5 & 7 \\ 0.1 & 0.01 & 0.001 & 0.0001 \end{bmatrix}$$

e sua transposta A^T*. Na sintaxe de* ***Julia*** *devemos escrever*

```
A=[11 34 6.7 4.9;
1 1 1 1;
2 4 5 7;
0.1 0.01 0.001 0.0001]
```

A transposta de A se indica simplesmente com A'*. Então o comando A*A' fornece o produto que é a matriz*

$$AA^T = \begin{bmatrix} 1345.9 & 56.6 & 225.8 & 1.44719 \\ 56.6 & 4.0 & 18.0 & 0.1111 \\ 225.8 & 18.0 & 94.0 & 0.2457 \\ 1.44719 & 0.1111 & 0.2457 & 0.010101 \end{bmatrix}$$

18.1 Caminhos num grafo

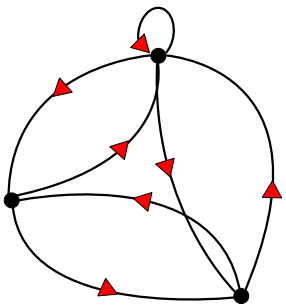

Figura 18.1: Grafo com 3 vértices e 7 arestas.

Vamos agora aplicar o conceito de produto de matrizes a um problema em Teoria de Grafos. Para ilustrar, vamos começar considerando um exemplo bem específico. Considere o grafo dado na Figura 18.1.

Problema 4. *De quantas formas diferentes é possível ir do vértice 2 ao vértice 3 fazendo-se exatamente 3 travessias?*

Nesse caso, o problema pode ser resolvido por inspeção direta: existem 4 formas diferentes de ir de 2 até 3 fazendo-se exatamente 3 travessias.

Vamos agora olhar para a matriz de adjacência do grafo (veja a Definição 11.1.1). Ela é $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$. Considere agora as potências da matriz de adjacência,

$$A^2 \stackrel{\text{def}}{=} AA = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 2 & 1 \\ 2 & 1 & 2 \end{bmatrix}$$

e

$$A^3 \stackrel{\text{def}}{=} A^2A = \begin{bmatrix} 7 & 5 & 5 \\ 5 & 3 & 4 \\ 5 & \mathbf{4} & 3 \end{bmatrix}.$$

Note que a entrada 32 da matriz A^3 é exatamente 4, a resposta do problema. Isso é coincidência? Se você parar para pensar, cada linha da i matriz A te diz a partir de quais vértices j você consegue chegar no vértice i. O que acontece, por exemplo, quando você faz o produto interno $\langle \ell_1, c_3 \rangle$ entre a primeira linha e a terceira coluna? Ora, cada entrada na terceira coluna te diz: partindo-se do vértice 3 para quais vértices posso ir! No caso aqui, podemos ir para 1 e 2. E como a primeira linha te diz de quais vértices você chega em 1, no produto interno $\langle \ell_1, c_3 \rangle$ só vamos ter parcelas 1×1 quando for possível sair de 3, passar por algum vértice e chegar no 1. E isso pode ser feito apenas de duas formas diferentes, que é igual a entrada 12 da matriz A^2. Podemos generalizar esse raciocínio e enunciar o seguinte teorema.

Teorema 18.1.1. *Seja $\mathcal{G}$ um grafo com n vértices e seja $A = [a_{ij}]_{n \times n}$ a matriz de adjacência de $\mathcal{G}$. Seja $k \in \mathbb{N}$ e considere k-ésima potência da matriz A, denotada por $A^k = [a_{ij}^k]_{n \times n}$.*

Então, o número de maneiras diferentes de se ir do vértice j ao vértice i no grafo fazendo-se exatamente k travessias é iguala a entrada a_{ij}^k da matriz A^k.

Para ilustrar a força desse resultado, vamos resolver uma versão bem mais difícil do problema anterior

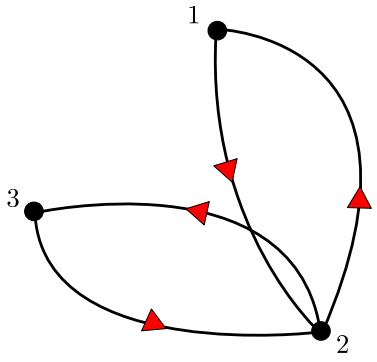

Figura 18.2: Caminhos em malta: o vértice 2 representa a ilha de Malta, o vértice 2 a ilha de Comino e o vértice 3 a ilha de Gozo.

Problema 5. *O arquipélago de Malta, no sul da Europa, é composto por três ilhas: Malta, Comino e Gozo. Existem barcos ligando Comino à Gozo e Comino à Malta. Determine de quantas formas diferentes é possível ir de Malta à Gozo fazendo-se exatamente* 666 *travessias.*

Solução. Considere o grafo descrito no enunciado que está representado na Figura 18.2. A matriz de adjacência é $A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$. Observe que $A^2 = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix}$ e $A^3 = \begin{bmatrix} 0 & 2 & 0 \\ 2 & 0 & 2 \\ 0 & 2 & 0 \end{bmatrix}$. Por indução podemos ver que as potências de A vão obedecer ao seguinte padrão: $A^{2n} = \begin{bmatrix} 2^{n-1} & 0 & 2^{n-1} \\ 0 & 2^n & 0 \\ 2^{n-1} & 0 & 2^{n-1} \end{bmatrix}$ e $A^{2n+1} = \begin{bmatrix} 0 & 2^n & 0 \\ 2^n & 0 & 2^n \\ 0 & 2^n & 0 \end{bmatrix}$. Como 666=2x333, temos então que $a_{31}^{666} = 2^{332}$. Portanto, aplicando o teorema 18.1.1 concluímos que existem 2^{332} maneiras diferentes de ir de Malta a Gozo fazendo-se exatamente 666 travessias. □

O problema dos caminhos em Malta evidencia como a solução matricial é capaz de revelar detalhes mais profundos do assunto. Por exemplo, por causa do padrão que têm as potências

da matriz A vemos que não é possível ir de Malta a Gozo fazendo-se um número ímpar de travessias e que o número de soluções do problema cresce exponencialmente quando aumentamos a quantidade de travessias possíveis. Compare com o problema anterior.

Vamos agora usar **Julia** e o Teorema 18.1.1 para resolver mais um problema.

Problema 6. *Suponha que A_1, A_2, A_3 e A_4 sejam páginas da* Internet. *Considere a matriz quadrada A na qual a entrada a_{ij} é 1 se a página A_i faz um* link *para a página A_j, e 0 caso contrário. Suponha que*

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}.$$

Para cada $i, j = 1, 2, 3, 4$ determine de quantas formas diferentes é possível ir da página A_i para a página A_j com, no máximo, 4 cliques.

Demonstração. Colocando A em Julia e fazendo-se o comando $A + A^2 + A^3 + A^4$ obtemos a matriz $B = \begin{bmatrix} 5 & 6 & 3 & 5 \\ 3 & 4 & 3 & 3 \\ 6 & 6 & 4 & 6 \\ 5 & 6 & 3 & 5 \end{bmatrix}$. □

18.2 Composição de funções lineares

Sejam $f : \mathbb{R}^d \to \mathbb{R}^k$ e $g : \mathbb{R}^n \to \mathbb{R}^d$ funções lineares. A composta de f e g é a função $f \circ g : \mathbb{R}^n \to \mathbb{R}^k$ definida por

$$f \circ g(x) \stackrel{\text{def}}{=} f(g(x)), \ \forall \ x \in \mathbb{R}^d.$$

Lema 18.2.1. *$f \circ g$ é linear*

Demonstração. Sejam $x, y \in \mathbb{R}^n$ e $\lambda \in \mathbb{R}$. Como g é linear

$$g(x + \lambda y) = g(x) + \lambda g(y).$$

Como f é linear,

$$f\left(g(x)+\lambda g(y)\right)=f(g(x)+\lambda f(g(y))=f\circ g(x)+\lambda f\circ g(y).$$

Portanto

$$f\circ g(x+\lambda y)=f\left(g(x+\lambda y)\right)=f\circ g(x)+\lambda f\circ g(y).$$

Isso estabelece a linearidade de $f\circ g$. □

Proposição 18.2.2. *Sejam $F=[a_{ij}]_{k\times d}$ a matriz de f na base canônica e $G=[b_{ij}]_{d\times n}$ na base canônica. Então a matriz de $f\circ g:\mathbb{R}^n\to\mathbb{R}^k$ é o produto FG.*

Demonstração. Seja $A=[c_{ij}]_{k\times n}$ a matriz de $f\circ g$ na base canônica. Então as colunas de A são os vetores

$$c_1^A=f\circ g(e_1),...,c_n^A=f\circ g(e_n).$$

Mas, se denotarmos por c_j^G a j-ésima coluna da matriz G teremos

$$f\circ g(e_j)=f(g(e_j))=f(c_j^G)=\begin{bmatrix}\langle\ell_1^F,c_j^G\rangle\\ \vdots\\ \langle\ell_k^F,c_j^G\rangle\end{bmatrix}.$$

Portanto,

$$c_j^A=\begin{bmatrix}\langle\ell_1^F,c_j^G\rangle\\ \vdots\\ \langle\ell_k^F,c_j^G\rangle\end{bmatrix}.$$

Pela definição do produto de matrizes isso prova que $A=FG$. □

Capítulo 19

Matriz inversa

Seja $f : \mathbb{R}^d \to \mathbb{R}^d$ linear. Dizemos que f é *invertível* se existe $g : \mathbb{R}^r \to \mathbb{R}^d$ tal que

$$f \circ g(x) = g \circ f(x) = x, \quad \forall \ \ x \in \mathbb{R}^d.$$

Ou seja, $f \circ g = Id$ onde $Id : \mathbb{R}^d \to \mathbb{R}^d$ é a função identidade, ou seja, $Id(x) = x$.

Observação 19.0.1. *Como $Id(e_j) = e_j$ temos que a matriz de Id é*

$$Id = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}.$$

Portanto, se F é a matriz de f e G é a matriz da inversa de f, temos $FG = Id$.

Observação 19.0.2. *$f : \mathbb{R}^d \to \mathbb{R}^d$ é invertível se, e somente se, f é injetiva e sobrejetiva ao mesmo tempo. Com efeito, se f é invertível com $g : \mathbb{R}^d \to \mathbb{R}^d$ a inversa de f então $y = g(x)$ é a única solução da equação $f(y) = x$. Logo f deve ser sobrejetiva e injetiva. Reciprocamente, se f é injetiva e sobrejetiva ao mesmo tempo então, para cada $x \in \mathbb{R}^d$, a equação $f(y) = x$ possui uma única solução. Defina $g(x) = y$ como sendo essa única solução. A função assim definida "herda"a linearidade de f:*

$$g(x^1 + \lambda x^2) = g(f(y^1) + \lambda f(y^2)) = g(f(y^1 + \lambda y^2)) = y^1 + \lambda y^2 = g(x^1) + \lambda g(x^2).$$

A observação acima motiva o

Lema 19.0.3. *f é injetiva se, e somente se,* $\dim N(f) = 0$

Demonstração. Suponha f injetiva. Então, como $f(0) = 0$ o núcleo de f tem que ser trivial. Reciprocamente, se $f(x) = f(y)$ então, pela linearidade de f, $f(x - y) = 0$ e portanto $x - y \in N(f)$. Portanto, se $\dim N(f) = 0$ então $x - y = 0$, ou seja, $x = y$. Isso prova que f é injetiva e estabelece o resultado. □

Teorema 19.0.4. *Seja $f : \mathbb{R}^d \to \mathbb{R}^d$ linear. Então f é invertível se, e somente se,* $\dim N(f) = 0$.

Demonstração. Pelo Teorema do Núcleo e da Imagem, $\dim N(f) = 0 \iff \dim Im(f) = d$. Além disso, f é sobrejetiva se, e somente se $\dim Im(f) = d$. Ou seja, a injetividade de f é equivalente a sua sobrejetividade. Portanto juntando o lema e a observação concluímos. □

Capítulo 20

Determinantes: significado e propriedades

Capítulo 21

O método dos mínimos quadrados

Capítulo 22

Autovalores e autovetores

Capítulo 23

Potência de matrizes

Capítulo 24

O Teorema Espectral